HEALTHINFO 2021

The Sixth International Conference on Informatics and Assistive Technologies for Health-Care, Medical Support and Wellbeing

ISBN: 978-1-61208-916-4

October 3 -7, 2021

Barcelona, Spain

HEALTHINFO 2021 Editors

Andrea Corradini, Copenhagen School of Design and Technology, Denmark

Jesus Zegarra Flores, Capgemini Engineering, France

HEALTHINFO 2021

Forward

The Sixth International Conference on Informatics and Assistive Technologies for Health-Care, Medical Support and Wellbeing (HEALTHINFO 2021), held on October 3 - 7, 2021 in Barcelona, Spain, tackles with particular aspects belonging to health informatics systems, health information, health informatics data, health informatics technologies, clinical practice and training, and wellbeing informatics in terms of existing and needed solutions.

The progress in society and technology regarding the application of systems approaches information and data processing principles, modeling and information technology, computation and communications solutions led to a substantial improvement of problems in assistive healthcare, public health, and the everyday wellbeing. While achievements are tangible, open issues related to global acceptance, costs models, personalized services, record privacy, and real-time medical actions for citizens' wellbeing are still under scrutiny.

We take here the opportunity to warmly thank all the members of the HEALTHINFO 2021 technical program committee as well as the numerous reviewers. The creation of such a broad and high quality conference program would not have been possible without their involvement. We also kindly thank all the authors that dedicated much of their time and efforts to contribute to the HEALTHINFO 2021. We truly believe that thanks to all these efforts, the final conference program consists of top quality contributions.

This event could also not have been a reality without the support of many individuals, organizations and sponsors. We also gratefully thank the members of the HEALTHINFO 2021 organizing committee for their help in handling the logistics and for their work that is making this professional meeting a success.

We hope the HEALTHINFO 2021 was a successful international forum for the exchange of ideas and results between academia and industry and to promote further progress in health informatics research.

HEALTHINFO 2021 Steering Committee

Tomohiro Kuroda, Kyoto University / Kyoto University Hospital, Japan
Nelson P. Rocha, University of Aveiro, Portugal

HEALTHINFO 2021 Publicity Chair

José Miguel Jiménez, Universitat Politecnica de Valencia, Spain
Lorena Parra, Universitat Politecnica de Valencia, Spain

HEALTHINFO 2021

Committee

HEALTHINFO 2021 Steering Committee

Tomohiro Kuroda, Kyoto University / Kyoto University Hospital, Japan
Nelson P. Rocha, University of Aveiro, Portugal

HEALTHINFO 2021 Publicity Chair

José Miguel Jiménez, Universitat Politecnica de Valencia, Spain
Lorena Parra, Universitat Politecnica de Valencia, Spain

HEALTHINFO 2021 Technical Program Committee

Djafar Ould Abdeslam, University of Haute Alsace, France
Faris M. Abomelha, King Faisal Specialist Hospital & Research Centre, Kingdom of Saudi Arabia
Miriam Allalouf, Azrieli College of Engineering Jerusalem - JCE, Israel
Jens Allmer, Hochschule Ruhr West, University of Applied Sciences, Germany
João R. Almeida, University of Aveiro, Portugal / University of A Coruña, Spain
Shada Alsalamah, King Saud University, Saudi Arabia
Khalfalla Awedat, Pacific LutheranUniversity, USA
Mana Azarm, University of Ottawa, Canada
Panagiotis D. Bamidis, Aristotle University of Thessaloniki, Greece
Hugo Barbosa, Lusofona University of Porto / Faculty of Engineering of the University of Porto, Portugal
Fabio Baselice, University of Naples Parthenope, Italy
Arriel Benis, Holon Institute of Technology, Israel
Ahmed Bentajer, National School of Applied Sciences | Abdelmalek Essaad University, Tetouan, Morocco
Vilmos Bilicki, University of Szeged, Hungary
Amine Boufaied, ISITCom | University of Sousse, Tunisia
Guillaume Bouleux, University of Saint Etienne | INSA-Lyon, France
Klaus Brinker, Hamm-Lippstadt University of Applied Sciences, Germany
Tolga Çakmak, Hacettepe University, Turkey
Manuel Campos Martínez, University of Murcia, Spain
Rui Pedro Charters Lopes Rijo, Polytechnic of Leiria | INESCC | CINTESIS, Portugal
K.A.D. Chathurangika P. Kahandawaarachchi,Sri Lanka Institute of Information Technology, Sri Lanka
Ayan Chatterjee, University of Agder, Grimstad, Norway
Bhargava Chinni, University of Rochester, USA
Giulia Cisotto, University of Padova, Italy / National Centre for Neurology and Psychiatry of Tokyo, Japan
Alberto Cliquet Jr., UNICAMP / USP, Brazil
Zhou Congcong, Zhejiang University, China
Andrea Corradini, Copenhagen School of Design and Technology, Denmark
Giuseppe De Pietro, Institute for High Performance Computing and Networking (ICAR) - Italian National Research Council (CNR), Italy / Temple University's College of Science and Technology, Philadelphia, USA

Copyright Information

For your reference, this is the text governing the copyright release for material published by IARIA.

The copyright release is a transfer of publication rights, which allows IARIA and its partners to drive the dissemination of the published material. This allows IARIA to give articles increased visibility via distribution, inclusion in libraries, and arrangements for submission to indexes.

I, the undersigned, declare that the article is original, and that I represent the authors of this article in the copyright release matters. If this work has been done as work-for-hire, I have obtained all necessary clearances to execute a copyright release. I hereby irrevocably transfer exclusive copyright for this material to IARIA. I give IARIA permission or reproduce the work in any media format such as, but not limited to, print, digital, or electronic. I give IARIA permission to distribute the materials without restriction to any institutions or individuals. I give IARIA permission to submit the work for inclusion in article repositories as IARIA sees fit.

I, the undersigned, declare that to the best of my knowledge, the article is does not contain libelous or otherwise unlawful contents or invading the right of privacy or infringing on a proprietary right.

Following the copyright release, any circulated version of the article must bear the copyright notice and any header and footer information that IARIA applies to the published article.

IARIA grants royalty-free permission to the authors to disseminate the work, under the above provisions, for any academic, commercial, or industrial use. IARIA grants royalty-free permission to any individuals or institutions to make the article available electronically, online, or in print.

IARIA acknowledges that rights to any algorithm, process, procedure, apparatus, or articles of manufacture remain with the authors and their employers.

I, the undersigned, understand that IARIA will not be liable, in contract, tort (including, without limitation, negligence), pre-contract or other representations (other than fraudulent misrepresentations) or otherwise in connection with the publication of my work.

Exception to the above is made for work-for-hire performed while employed by the government. In that case, copyright to the material remains with the said government. The rightful owners (authors and government entity) grant unlimited and unrestricted permission to IARIA, IARIA's contractors, and IARIA's partners to further distribute the work.

Table of Contents

Early-Stage Epidemic Forecasting

Confirmation of the Grey Model Theory to Early Stage Epidemic Forecasting

Olayemi Olabisi
University of Liverpool
Liverpool, England
e-mail: olayemi_olabisi@hotmail.co.uk

Andrea Corradini
Copenhagen School of Design and Technology
Copenhagen, Denmark
e-mail: andc@kea.dk

***Abstract*—Statistical methods and machine learning methods are currently the most popular ways for forecasting. Some of these include autoregressive models, cumulative sum charts, growth models, Support Vector Machines for regression (SVR), Polynomial Neural Networks (PNNs) as well as several others. The inherent limitation associated with these models is that they require large sample size for accurate forecasts. In this paper, we investigate the ability to forecast a disease outbreak when data is limited by employing variations of the Grey Model (GM) forecasting. Lack or limitation of data is rather common in the early stages of a disease outbreak. We present the results of a simulation that shows the model's ability to leverage the exponential growth associated with the rate of spread of diseases in the early stages. A comparative analysis of our quantitative results using the coronavirus dataset of a few countries indicate that both the Gompertz model and the PNN model perform better than the traditional Grey Model GM(1,1) in fitting and forecasting. At the same time, qualitative evidence indicate that the Grey Model is suitable for early-stage epidemics provided methods of enhancement are employed i.e. improved background value calculations and better methods for accumulative generating operation as seen in the Fractional Grey Model FGM(1,1) algorithm.**

Keywords- Forecasting; Grey Forecasting model; Early Stage Epidemic; Coronavirus.

I. INTRODUCTION

On the 31st of December 2019, World Health Organization (WHO)'s country office in the people's republic of China picked up a media statement by the Wuhan municipal health commission from their website on cases of viral pneumonia in Wuhan. On the 1st of January 2020, the WHO activated its incident management support team (IMST) as part of its emergency response framework. On the 9th of January 2020, the WHO reported that the Chinese authorities have confirmed the outbreak of a novel coronavirus. It took eleven days for the confirmation of the outbreak of the virus. Since the official declaration of the novel coronavirus as a global pandemic, the world has witnessed an immense spread of the infection. By early September 2021, the virus has spread to nearly every country causing more than 215 million infection cases and nearly 4.5 million deaths worldwide [1], upending lives and derailing the global economy.

Accurate forecasting is one of the tools that are typically used to plan and mitigate risk after and during the outbreak of a disease. Its importance is something we have come to appreciate in the recent coronavirus outbreak. Unlike normal forecasting, epidemic forecasting exhibits traits which have been described in [2][3] especially when considering the initial period of the outbreak of a disease. This initial period can be identified by a sharp rise in cases where numbers are initially dormant. An example of this would be in cases of influenza, where at the beginning of the season little or no cases are observed, then a sudden surge in number of cases is observed [3].

There are several models, which can be employed for modelling and forecasting the initial stage of an epidemic. Examples of models used in previous studies include autoregressive integrated moving average, holt-winters multiplicative, linear regression, support vector machine for regression, fast decision learner, regression trees, PNNs and PNN+cf [4]. Other proposed models for forecasting disease outbreaks include generalized growth models as described in [5]. The problem inherent with these methods and with other statistical and machine learning techniques is that they require large datasets for higher levels of accuracy [6]. In the coronavirus case, it took eleven days for the virus to be identified as a threat and as such only eleven data points were available.

Small sample set problems [7][8] refer to the cases where the number of samples is less than 50 in respect to engineering applications or less than 30 regarding academic research. Such small number sample size cannot fully reveal the behavior of a system or population [9]. There are three main approaches for dealing with small samples forecasting. The first applicable method is the Grey forecasting model. This model uses accumulative generation operator (AGO) of the Grey theory to deal with raw sample set for improving the accuracy of forecasting [10]. A second method is the use of Virtual Sample generation (VSG) [10]. This latter method fills the information gaps among each raw sample to stabilize the forecasting performance through adding newly generated virtual samples. Some examples include mega trend diffusion (MTD) [11], generalized trend diffusion (GTD) [12], genetic algorithm based virtual sample generation (GAVSG) [13] and Gaussian distribution based on VSG [14]. A particular limitation of this approach was highlighted in a study on the data generated by Generative Adversarial Networks (GANS) that showed that Top-1 and Top-5 accuracy error increased by 120% and 384% respectively and the improvement on the classification accuracy was not significant [15]. Based on this result, it was concluded that the use of such technique may incur more damage than help especially in scenarios where

 ISBN: 978-1-61208-916-4

life might be on the line. A third approach for dealing with small sample size focuses on feature extraction. This involves dimension reduction whereby useful attributes or features are identified and selected to enhance the accuracy of the forecast. During the initial period of disease outbreaks, all information acquired should be used because information on the new disease is limited. The different factors which affect the spread of the disease might not be known and as such feature extraction does not seem to be a suitable approach [10].

Based on these findings, the aim of this paper is to investigate the applicability of the Grey forecasting algorithm (GM(1,1)) to the first take off period of the coronavirus outbreak. The rest of this paper is organized as follows. Section II describes shortly the state of the art in approaches to forecasting applied to epidemics. Section III outlines the sources for the data we used for out research along with the reason for such a choice and a short description of the evaluation criteria for the assessment of the performance of forecasting algorithms. Section IV describes a few common approaches to forecasting. Section V shows the result of our simulations, both in graphical and tabular forms, with different model and data from three different countries. Eventually, the conclusions close the article.

II. State of the Art

Most approaches to forecasting require the use of statistical or machine learning models which require large sample size for higher levels of accuracy [6]. This fact in conjunction with the fact that in the early stages of epidemics data is limited inspired the application of the Grey system theory to early-stage epidemic forecasting. A great deal of research has been successfully conducted to verify the application of the Grey system theory in case of epidemics. However, to our knowledge, very little research has been carried out to compare the performance of the Grey models to other existing models like PNNs or Growth models. In this paper, we aim to shed more light on the Grey forecasting model in hope to better equip us to fight against future epidemics.

III. Methodology

In this section, we outline the sources for the data we used for out research along with the reason for such a choice and a short description of the evaluation criteria for the assessment of the performance of forecasting algorithms.

A. Data Source

Due to the novelty and the united approach society undertook against the novel coronavirus (Covid-19) pandemic, there was an ample amount of data sources. In our study, we rely on secondary data sources resulting in time saving from data collection and aggregation. We considered various options that were referred to by reliable sources such as the WHO, the European Centre for Disease, and the John Hopkins University. We decided to not use WHO data as they had shifted the timing on their reports causing overlaps in their data timelines. The data from the European Centre for Disease (ECDC) and John Hopkins data was fairly similar instead. Eventually we chose the data available on the ECDC because it included a few pre-calculated variables which made the dataset more complete for out study. The ECDC data is available on ourworldindata.org. The website and the data were compiled with the mindset that approaching a common problem together is the best way to solve it. The website confirms the daily numbers of Covid-19 cases and provides various visualizations [16].

B. Sample

A sample size of 10 was chosen to reflect more realistic situations. The WHO announced the virus's potential after eleven days compared with only 6 – 8 data points used in other studies [4][17]. We chose the data from three specific countries after considering various factors like political influences, infrastructure capabilities, corruption, and coronavirus timings. Upon thorough investigation countries which were infected early by coronavirus appear to have been caught unprepared and so the data observed from such countries appear inconsistent and un-reliable. The three countries whose data we chose to use in our study are Albania, Haiti, and Argentina. These countries were severely hit by the virus after March 2021, that means at a time when most unaffected countries were expecting and preparing for the arrival of Covid-19. Ultimately, this resulted in ensuring an adequate reporting of cases.

C. Evaluation Criteria

To assess the performance of the forecasting algorithms, there are various criterions applicable such as the Mean Average Error (MAE), the Root Mean Square Error (RMSE), and the Mean Absolute Percentage Error (MAPE). We chose to use the MAPE because of its more intuitive interpretation.

D. Systematic Review

We carried out a systematic review employing the PRISMA guidelines [18] to further understand the different components of the Grey forecasting algorithm, how to optimize these components to increase the forecasting accuracy of the Grey forecasting model.

IV. Models

In this section, we provide an overview of a few common approaches to forecasting, including the Grey system model theory, Polynomial Neural Networks, and the Gompertz model.

A. Grey System Theory

The Grey system theory works on systems with partially identified and partially unknown information by drawing out crucial information through the generation and development of the partially known information. It can describe correctly and effectively the systemic operational behavior of many

 ISBN: 978-1-61208-916-4

systems such as social, economic, agricultural, industrial, ecological, and biological systems [19]. The Grey system theory uncovers laws of change by mining and structuring available raw data, thereby representing an approach of finding data out of data. This approach is referred to as Grey sequence generation. This method considers that even though the expression of an independent system might be complex, its data is chaotic. There must be some internal laws ruling the existence of the system and its operation. Hidden laws are found through the generation of Grey numbers or functions of sequence operator. These operators allow the Grey system to handle various attributes that may be present within the data to adequately extract the information embedded within the sequence. Some of these operators include buffer operators (e.g., weakening operator, strengthening operator), average generation operators, stepwise ratio generation operators, inverse accumulating generation operators and accumulating generation operators [20].

In most instances where the GM(1,1) algorithm has been applied in research, it has almost always been an enhanced version. The most common of enhanced parameters are background values and accumulative generative operators. Background values theoretically have two primary functions. The first function is to smooth the data and strip it of its randomness and the second is to emphasize the importance of the newest datum [19]. The precision and the prediction power of the Grey model depends on the accuracy of a few coefficients (called the a and b coefficients), whereby the values of these parameters depend on the original data sequence and the background value, again reflecting the importance of the background value. Hence by improving the background value, the a and b coefficients are also improved having a compounding effect on the accuracy and performance of the model. The traditional GM(1,1) models background value is always a default value of 0.5 which is a high simplification of the modelling process meaning no preference is given to old or new data. More realistically, the value of the background value coefficient should be calculated based on the original dataset. The Accumulative Generative Operational algorithm (AGO) and the Inverse Cumulative Generative Operational algorithm (I-AGO) are important parts of the Grey model algorithm and are responsible for reducing the variation or fluctuations of the original series. The problem found in the algorithms used for the AGO sequence generation comes from the definition of class ratios when creating the Grey model algorithm by assuming that each component of the original sequence is positive and the class ratio of the first AGO manipulation is always smaller than the class ratio of the second AGO manipulation [20]. This means that in some cases the first accumulated generating operator violates the principle of new information priority and principle of minimal information of Grey system theory. However, it can be proved that the condition stated above is only sufficient but is not a necessary one [21].

B. Polynomial Neural Network

A previous study investigated the early-stage epidemic forecasting and concluded that the best algorithm is the Polynomial neural networks with corrected feedback with an RMSE of 136 [4]. PNNs belong to a group of neural algorithms whose theory roots back to and is based on works from the early 1970s [22]. Those types of neural networks are "self-organizing" networks to express the fact that the connections between the neurons in the network are selected during the training phase to optimize the network. The number of layers in the network are also determined automatically to produce maximum accuracy without overfitting [23].

C. Gompertz Model

Growth models are currently the most used methods for early-stage forecasting of disease outbreaks. Recently, these models were applied to twenty infectious disease outbreaks representing a range of transmission routes and a wide range of epidemic profiles. The outbreaks ranged from slow growth (sub-exponential) to fast growth (close to exponential). The proposed generalized growth model outperformed the other algorithms that it was compared to [6].

Growth models can be grouped into at least two main categories. The first group includes models without inflection points such as the Brody and negative models [24]. The second group include models with sigmoidal shape that also poses a fixed inflection points such as the Gompertz, the logistic or the von Bertalanffy models [24]. The Gompertz and the logistic models are the most frequently adopted models in literature. Given that the cumulative number of cases by coronavirus appears to form an asymmetrical sigmoidal curve, the logistic and the Gompertz model naturally lend themselves as ideal candidates to apply [24].

The Gompertz model and logistic model are both growth models with similar properties, making them useful for the representation of the generalized growth models. There has not been a distinct advantage found between both models as both require constants corresponding to upper asymptote, time origin, and time unit or rate [25]. But in practice, it has been found that the logistic model gives good fits on material showing an inflection about midway between the asymptotes, while the inflection point for the Gompertz model is about 37%. This makes the Gompertz model more suitable for first take off period forecasting [25].

D. Average Ensemble

Enhancing the traditional GM(1,1) algorithm can be achieved in multiple ways, all of which offer different advantages and consequently have different effects on the algorithm. Based on the systematic review we conducted, the use of multi-model approach (ensembles) turned out to be one of the most popular methods and proved to be the most suitable method to apply in this paper. There are various types of ensembles methods namely bagging, boosting, Adaboost, random forest, gradient boosting, averages etc. [26].

 ISBN: 978-1-61208-916-4

For the purpose of this paper, we applied the average ensemble to the Covid-19 dataset in order to demonstrate the application of one of the methods of enhancement. Using the average of forecasts is one of the simple but effective methods of ensembles. Previous studies [27] suggest that using averages of forecasts provides improved forecasting accuracy and that the variability of accuracy among different combinations decreases, as the number of methods in the average increases. Similar earlier studies also indicate that, in some cases, the simple average ensemble can outperform its weighted average counterpart [28].

In the next section, we show the forecasting results we obtained using the coronavirus dataset of Albania, Argentina, and Haiti. The biggest influence on the countries chosen is the quality of data generated. Upon thorough observations of the data, countries like China who were hit early by the virus seem to have been un-prepared meaning the necessary actions required to adequately record and publish the coronavirus numbers were not in place and so countries which contracted coronavirus after March 2020 were chosen as their published data appear more reliable and consistent.

V. RESULTS

In this section we show the result of our simulations with different model for a few selected countries.

A. Albania Forecast

Figure 1 shows how well each model describes the Albania dataset and forecasts the Covid-19 cases. The y axis shows the number of cases while the x axis shows the days that have elapsed. The red line shows the actual data. The other lines show the fit and forecast for the models we chose to employ, namely the GM(1,1), the Gompertz, the PNN, and the ensemble model. The graphics indicates that the best models for the Albania dataset are the Gompertz model and the PNN while the GM(1,1) appears to be the worst fitting one. Table I shows the MAPE values for all models we considered in out simulations when the data for Argentina was used.

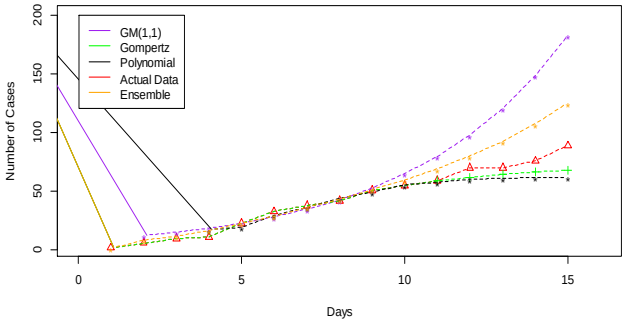


Figure 1. Model values for GM(1,1), Gompertz, PNN, Ensemble model for the Albania Dataset.

TABLE I. MAPE VALUES FOR ALL MODELS AND THE DATA SET FOR ALBANIA.

	In Sample	*Out of Sample*
GM(1,1)	30.67	69.76
Gompertz	10.50	11.16
PNN	13.16	15.53
Ensemble	15.92	29.35

B. Argentina Forecast

Figure 2 visually highlights the performance of the four different models applied to the dataset for the country Argentina. The y axis shows the number of cases while the x axis shows the days that have elapsed. Again, the red line shows the actual data available for Argentina. The GM(1,1) seems to forecast very high values, which is clear when observing the purple and the red line. The best performing model in this case appears indistinguishable but this can be explained by comparing the MAPE values as seen in Table II.

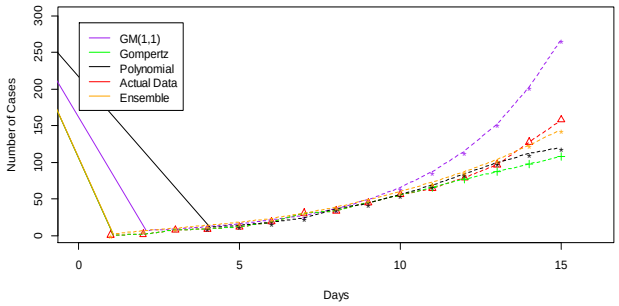


Figure 2. Forecast values for GM(1,1), Gompertz, PNN, Ensemble model for the Argentina Dataset.

TABLE II. MAPE VALUES FOR ALL MODELS AND THE DATA SET FOR ARGENTINA.

	In Sample	*Out of Sample*
GM(1,1)	43.28	53.11
Gompertz	18.69	13.90
PNN	12.24	10.27
Ensemble	54.67	8.47

C. Haiti Forecast

Figure 3 highlights visually how well the models fit and forecasts the Haiti dataset. In this case, the Gompertz model appears to have a particularly good fit for the data set, but aside from this, it can be difficult to distinguish and discriminate using this graph only. Table III outlines the MAPE values and thus can provide a clearer understanding on model accuracy for the data related to Haiti.

 ISBN: 978-1-61208-916-4

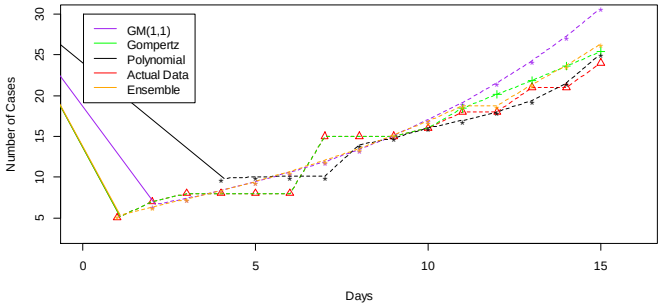


Figure 3. Forecast values for GM(1,1), Gompertz, PNN, Ensemble model for the Haiti Dataset.

TABLE III. MAPE VALUES FOR ALL MODELS AND THE DATA SET OF HAITI.

	In Sample	*Out of Sample*
GM(1,1)	11.73	19.90
Gompertz	11.83	7.33
PNN	16.67	4.15
Ensemble	11.43	6.53

As already mentioned, the MAPE values give a better intuitive understanding of forecast accuracy. Overall, the Gompertz models and the PNN both perform equally well on different datasets suggesting the need for further tests using other datasets. The GM(1,1) consistently performed as the worst. While the ensemble was implemented to show the potential for improving the GM(1,1), the improvement can be attributed to the outstanding performance of the Gompertz model.

D. Systematic Review

These results from the experiments initiated an investigation into the methods which the Grey forecasting model's parameters could be optimized to provide better accuracy. From the studies that we reviewed systematically, 38% improved the background value of the Grey forecasting model, 23% used a multi-model approach similar to an ensemble while 15% enhanced the Grey forecasting model GM(1,1) by improving the accumulative generating operation mechanism, other methods employed including the use of Virtual sample generation (VSG), parameter optimization, and finally the diffusion models.

VI. CONCLUSION

The comparison of the GM(1,1) forecasting algorithm to the PNN and Gompertz model against the datasets showed that the traditional GM(1,1) algorithm was not able to match the forecasting capabilities of the Gompertz and PNN model. This result was unexpected but has been attributed to the sub-exponential nature of the dataset and the hidden nonoptimized parameters within the GM(1,1) algorithm. While we see that the enhanced ensemble GM(1,1) algorithm performs better in comparison to the traditional GM(1,1), enhanced GM(1,1) models which optimize the parameters within the GM(1,1) model itself have immense potential in the forecasting of early stage epidemic forecasting.

The preliminary conclusions of this paper are confirmed in a separate research study by Liu and colleagues [29]. Liu uses fractional Grey model FGM(1,1) that is an enhanced version of the traditional forecasting model GM(1,1) by using a fractional order accumulation. Liu's FGM(1,1) model can better adhere to the principle of new information priority and was applied to forecast the first take off period of Covid-19 with good results. In another study [30], the GM(1,1) model was also applied to and compared with other modifications such as the non-linear Grey Bernoulli model (NGBM(1,1) and the FGM(1,1) model to long term cumulative forecasting of the Covid-19 epidemic cases in UK, USA and Italy. The values of root mean square error, absolute percentage error (APE) and R^2 values for GM (1,1), NGBM (1,1) and fractional accumulative nonlinear Grey Bernoulli model (FANGBM(1,1)) for the prediction of the cumulative cases of Covid-19 indicate that the FANGBM is the most accurate with the highest R^2 and lowest RMSE and APE values.

Based on all this evidence, the aim of this paper to confirm the applicability of Grey forecasting model to forecast early-stage epidemic cases has been achieved.

REFERENCES

[1] World Health Organisation (WHO) (2020) Coronavirus disease Weekly Epidemiological Update and Weekly Operational Update https://www.who.int/emergencies/diseases/novel-coronavirus-2019/ [Accessed 02/09/2020].

[2] Unkel, F., Farrington, P., & Garthwaite, P., (2012) Statistical Methods for the Prospective Detection of Infectious Disease Outbreaks: a review. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, Vol.175, Nr. 1, pp. 49-82.

[3] Tabataba, F.S, Prithwish, C., Ramakrishman, N., Venkatramanan, S., Jiangzhou, C., Lewis, B., & Marathe, M. (2017) A Framework for evaluating epidemic forecasts, *BMC Infectious Diseases* Vol. 17, Nr. 345.

[4] Fong, S.J., Li, G., Dey, N., Crepo, R.G., & Herrera-Viedma, E. (2020) Finding an Accurate Early Forecasting model from small dataset: A case of 2019-ncov novel coronavirus outbreak. *International Journal of Interactive Multimedia and Artificial Intelligence*, Vol. 6. Nr. 1, pp. 132-40

[5] Viboud, C., Simonsen, L., & Chowell, G., (2016) A generalised growth model to characterise the early ascending phase of infectious disease. *Epidemics*, Vol. 15, June 2016, pp 27-37.

[6] Hyndman, R.J., & Athanasopoulos, G. (2018) *Forecasting: principles and practice*, 2nd edition, OTexts: Melbourne, Australia. OTexts.com/fpp2. [Accessed 02/09/2020]

[7] Zhu, B., Chen, Z., & Le, Y., (2016). A novel mega-trend-diffusion for small sample. *CIESC Journal*, Vol. 67, pp. 820-826.

[8] Chang, C.J., Li, D.C., Dai, W.L., & Chen, C.C. (2014) A Latent function to extend domain attributes to improve the accuracy of small data set forecasting. *Neurocomputing*, Vol. 129, April 2014, pp. 343–349.

[9] Li, D.C., Fang, Y.H., Lai, Y.Y., & Hu, S.C. (2009) Utilization of virtual sample to facilitate cancer identification for DNA microarray data in the early stages of an investigation. *Information Sciences*, Vol. 179, Nr. 16, pp. 2740-2753.

 ISBN: 978-1-61208-916-4

[10] Chen, Z., Zhu, B., He, Y., & Yu, L. (2017) A PSO based virtual sample generation method for small sample sets: Applications to regression datasets. *Engineering Applications of Artificial Intelligence*, Vol. 59, March 2017, Pages 236-243.

[11] Li, D.C., Wu, C.S., Sai, T.I., & Lina, Y.S. (2007) Using mega-trend diffusion and artificial samples in small dataset for early flexible manufacturing system scheduling knowledge. *Computers & Operations Research*, Vol. 34, Nr. 4, pp. 966-982.

[12] Lin, Y.S., & Li, D.C. (2010) The generalization trend-diffusion modeling algorithm for small datasets in the early stages of manufacturing systems. *European Journal of Operational Research* Vol. 207, Nr. 1, pp 121-130.

[13] Li, D.C., & Wen, I.H. (2014) A genetic based virtual sample generation technique to improve small data set learning. *Neurocomputing*, Vol. 143, Nr., pp. 222-230.

[14] Yang, J., Yu, X., Xie, Z., & Zhang, J. (2011) A novel virtual sample generation method based on Gaussian distribution. *Knowledge-Based Systems*, Vol. 24, Nr. 6, pp. 740-748.

[15] Ravuri, S., & Vinyals, O. (2019) Seeing is not necessarily Believing : Limitations of BigGans for Data Augmentation.

[16] Ritchie, H., Mathieu, E., Rodés-Guirao, L., Appel, C., Giattino, C. Ortiz-Ospina, E., Hasell, J., Macdonald, B., Beltekian, D., & Roser, M. (2020) Coronavirus Pandemic (COVID-19). Published online at: https://ourworldindata.org/coronavirus [Accessed 02/09/2020]

[17] Wu, K., Darcet, D., Wang, Q., & Sornette, D. (2020) Generalised Logistic growth modeling of the COVID-19 outbreak in 29 provinces in China and in the rest of the world. *Nonlinear Dynamics*, Vol. 101, pp. 1561–1581.

[18] Page, M.J., McKenzie, J.E., Bossuyt, P.M., Boutron, I., Hoffmann, T.C., & Mulrow, C.D. (2020). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*, March 2021, Nr. 372.

[19] Lin, Y., & Liu, S. (2004) A Historical Introduction to Grey systems Theory. In: Proceedings of IEEE International Conference on Systems, Man and Cybernetics, Vol.3, pp. 2403-2408.

[20] Deng, J.L. (1982) Control problems of Grey systems, *Systems & Control Letters*, Vol. 1 No. 5, pp. 288-94.

[21] Wen, K.-L., Dai, J.-W., Chang, T.-C., Tong, C-C. (2001) The discussions of class ratio for AGO algorithm in Grey theory. In: Proceedings of IEEE International Conference on Systems, Man and Cybernetics, Vol. 1, pp. 7-11.

[22] Ivakhnenko, A.G. (1971) Ploynomial Theory of Complex Systems. IEEE Transactions on Systams, Man, and cybernetics, Vol. 1, Nr. 4, pp. 364-378.

[23] Dag, O., Yozgatligil, C. (2016) GMDH: An R Package for Short-Term Forecasting via GMDH-Type Neural Network Algorithms. *The R Journal*, Vol. 8, Nr. 1, pp. 379-386.

[24] Mandujano Valle, J.A. (2020) Predicting the number of total COVID-19 cases and deaths in Brazil by the Gompertz model. *Nonlinear Dynamics*, Vol. 102, pp. 2951–2957.

[25] Winsor, C.O. (1931) The Gompertz Curve as a Growth Curve. *Proceedings of the Natlional Academy of Science USA*, Vol. 18, Nr. 1, pp. 1–8.

[26] Latka, M., Ignaccolo, M., Jernajczyk, W., & West, B. (2010). Time and ensemble averaging in time series analysis.

[27] Makridakis, S., & Winkler, R. L. (1983) Averages of Forecasts - Some Empirical Results, *Management Science*, Vol. 29, pp. 987-996.

[28] Palm, F. C. , & Zellner, A. (1992) To combine or not to combine - Issues of combining forecasts, *Journal of Forecasting*, vol. 11, pp. 687-701.

[29] Liu, L., Chen, Y,. & Wu, L. (2020) Forecasting Confirmed Cases, Deaths, and Recoveries from COVID-19 in China during the Early Stage. Mathematical Problems in Engineering. 1-4. 10.

[30] Şahin, U. & Şahin, T. (2020) Forecasting the cumulative number of confirmed cases of COVID-19 in Italy, UK and USA using fractional nonlinear Grey Bernoulli model. *Chaos Solitons & Fractals*, Vol. 138.

 ISBN: 978-1-61208-916-4

A Goal-Driven Framework for Individualised Self-Care for Early-Stage Dementia

Jonathan Turner, Ciarán Nugent, Damon Berry, Dympna O'Sullivan
Department of Computer Science
Technological University Dublin
Dublin, Ireland
email: jonathan.turner@tudublin.ie, ciaran.nugent@tudublin.ie, damon.berry@tudublin.ie, dympna.osullivan@tudublin.ie

Julie Doyle, Michael Wilson, Ann Marron
NetwellCASALA
Dundalk Institute of Technology
Dundalk, Ireland
email: julie.doyle@dkit.ie, michael.wilson@dkit.ie, ann.marron@hse.ie

***Abstract*— For people with early or moderate dementia, there are benefits to them continuing to live in their own homes for as long as possible, both in improved quality of life and associated measures such as increased social contact, increased physical activity and lower use of medication, and reduced costs and burden of care. Tools to help extend period of independent living, and to maintain quality of life in this period, are lacking. Systems exist to monitor individuals for problems, e.g. falls or wandering from the home, but there is scope for development of computerised support to help maintain activity in independent living. We aim to monitor achievement of activities, by app and by sensors, and provide recommendations on how to best maintain activities.**

Keywords- dementia; goal modelling; self-care; independent living.

I. INTRODUCTION

Dementia is a set of symptoms that may include deterioration in memory and ability to focus attention, unpredictable behaviour and decline in the ability to perform routine activities of everyday living. It can be caused by a number of conditions, with Alzheimer's disease being the most common cause of dementia [1]. Dementia symptoms are progressive and the disorder is incurable; Persons Living With Dementia (PLWD) require increasing care as their underlying disease progresses. Currently, it is estimated that about 50 million people worldwide suffer from a form of dementia, with this number projected to rise to more than 131 million by 2050, reflecting aging populations around the world [2]. As well as the human cost of the disease, there is a societal financial burden, estimated to be over $1 trillion a year worldwide, with 80-85% due to paid social care services and unpaid informal family care [3]. For people with early or moderate dementia, there are benefits to them continuing to live in their own homes for as long as possible, both in improved quality of life and associated measures such as increased social contact, increased physical activity and lower use of medication, and reduced costs of care [4]. Lower physical function is associated with increased risk of admission to long-term care, and so it is beneficial to maintain physical function for as long as possible [4]. It should be noted that only a minority of PWLDs will move into a care home [5].

The Smart Dementia Care project aims to develop a computerised toolkit to assist people with early-stage dementia to live independently in their own homes. This toolkit will support a PLWD and their family carer(s) in developing a personalised care plan, encourage each PLWD's compliance with their care plan, and will incorporate goal targets derived from care plans, existing models of daily activities, and activities defined by the individual PLWDs and their carers. The project is following a co-design methodology where the research team includes PLWDs as experts in their condition. Achievement of goals by a PLWD will be measured by a combination of self- (or carer-) reporting via an app, and automatic data collection from body-worn or static sensors. The app and sensor data capture are under development in parallel. In this paper, we focus on the development of a computational goal framework to enable the specification of goals by PLWD and the capture and quantification of goal achievement data.

Persons with dementia should be at the centre of decision-making about their care including basic activities (e.g., feeding, dressing), advanced activities (e.g., finances, using transportation), and meaningful activities (e.g., social and recreational pastimes). However, there is limited involvement of people living with dementia in the design of technology to support their care. We describe the early development of a goal framework to enable personalisation of goals and tracking of goal achievement for persons with an early-stage dementing illness with the aim of promoting independent living at home. Goals are developed from existing measures of human ability and disability, from activities defined in care plans, and activities that individuals desire to maintain as pleasurable activities. We present an initial outline of our framework, that includes the capture of information pertaining to each goal in each individual's personalised goal schedule, comparison of achievement with their personal target for included activities, and suggested actions to take in the case of under-achievement. The rest of this paper is organized as follows. Section II describes the process of selection of activities to be monitored. Section III

 ISBN: 978-1-61208-916-4

describes the chosen activities. Section IV discusses possible burdens on PLWDs of this work. Section V gives conclusions and an akcnowlegement.

II. METHOD

We wish to encourage PLWD to achieve goals that will help maintain their ability to live independently in their own homes for as long as possible and which will help them retain those activities that will continue to benefit quality of life for those who may move from their own home into residential care.

In particular, we wish to encourage PWLDs to perform activities that fall into one of three groups:

(i) Basic Activities of Daily Living (ADLs) and Instrumental Activities of Daily Living (IADLs). These activities are the most basic activities that allow individuals some degree of independence, which for ADLs include ability to bathe oneself, select clothes and dress oneself in the correct sequence, toileting, transfer, continence and feeding [6] and for IADL are the more sophisticated instrumental activities such as shopping, meal preparation, laundry and money management [7].

(ii) Activities defined in their care plans. Care plans are agreed between health and care professionals, PLWD and their family carers. They are documents that record the care needs of a PLWD and their carers. Care plans may include information about accessing services, monitoring co-morbidities, a named healthcare contact, and details of activities that the PLWD should aim to continue. The plan may include goals which may be related to performing specific activities. These care plans are relatively static, often paper-based and reviewed annually, not computerised or interactive, and so the degree of responsive personalisation for each PLWD is limited.

(iii) Activities that bring pleasure or satisfaction to PLWD. Example sets of such pleasurable activities exist but it is unlikely that any one PLWD will wish to enjoy all activities in any one defined set, and they may have wishes to enjoy activities that are not in any existing sets of pleasurable activities. In proposing and selecting pleasurable activities, we take inspiration from work from such organisations as the NHS [8]

In order to achieve our aims of extending the period spent in independent living and maintaining the quality of life in independent living, ADLs and IADLs will be standard for all PLWD, but other activities will be personalised using the care plans and wishes of each individual PLWD. Monitoring changes in goal achievement over time allows for a greater degree of personalisation for each PLWD. We reviewed the literature relating to basic activities of daily living, including scales used in determining the abilities and disabilities of PLWD and any goals for these basic activities. Similarly for pleasurable activities we reviewed the literature relating to measuring the frequency and degree of such activities and of related emotions.

III. RESULTS

Goal-oriented Cognitive Rehabilitation (CR) is a form of therapy which aims to address and manage functional disability and maximise social participation and engagement using a person-centred, goal-oriented and problem-solving approach [9]. It uses evidence-based rehabilitative methods and involves PLWDs and their carers or family members working together with a therapist to identify meaningful and personally relevant goals related to everyday activities [10]. Strategies are collaboratively devised and implemented; evaluation in terms of progress towards goal attainment is based on both participant and informant-reported information [9][11]. A multicentre randomized controlled trial by Clare et al [9] demonstrated that PLWDs were able to identify goals they felt were important and were motivated to address and attempt to attain. In addition, results from their trial suggest that an individualised and goal-oriented implementation of CR can lead to improvements in everyday functioning, and it can be an effective intervention for people with early-stage dementia. One of the fundamental strengths identified from Clare et al's work was the possibility of transfer and generalisation, with the goals that were identified and worked towards being relevant and applicable to improved functioning in real-world situations. Moreover, the suggestion has been made that once delivered in a cost-effective manner, CR could be integrated into care pathways with a view to developing strategies for living with dementia in the community [9].

Phinney et al [12] identified four categories of goal, based on interviews with and observations of PLWDs: (1) leisure and recreation; (2) household chores; (3) social involvements; (4) work-related. Interview participants stressed the importance of being able to continue engaging in these activities and the willingness to employ new strategies to do so should this be required as the disease progresses. Related to this is the importance of identifying those activities that were valued and enjoyed before the onset of dementia since these are likely to be considered intrinsically meaningful in terms of everyday life and past experience [12].

The Goal Attainment Scale (GAS) [13] has been used to measure achievement of goals in dementia, but GAS itself is not specific for dementia and the particular goals used in any analysis are selected in cooperation with the PWLD. There are a number of scales used to measure individuals' performance and disability, e.g., [14], but we found only two scales that were specific for PLWDs: The Pleasant Events Schedule for Alzheimer's Disease (PES-AD) of Teri & Logsdon [15] and the Dementia Quality of Life (DEMQOL) questionnaire of Smith et al [16].

(i) Basic activities of daily living. The long-established ADL scale of Katz [6] and IADL scale of Lawton and Brody [7] measure the ability of individuals to perform basic activities of daily living. These scales were created to allow scoring of the degree of self-maintenance functions that individuals are able to perform for themselves, with the Katz scale measuring more basic functions than the Lawton-Brody scale. It is important for these activities to be maintained for as long as possible both for those with dementia and those without dementia, but for those with some forms of dementia (including Alzheimer's), once the ability to perform an activity is lost, it cannot be regained. In Ireland, the Modified

 ISBN: 978-1-61208-916-4

Winchester Disability Scale (MWDS) [17] dominates assessment of risk in the elderly. Additional to the ADL and IADL scales it includes mental and social activity. It should be noted that none of these scales are specific for PWLD. Example activities measured by these three scales include the ability to bathe oneself; ability to select clothes and dress oneself appropriately; the ability to make telephone calls; ability to shop for oneself; and mobility.

(ii) Activities in individuals' care plans. Formal Care Plans contain a variety of information useful to the PLWD and their family carers, and are usually drawn up in cooperation with health professionals. They can include such useful and essential information as medication and prescriptions details, emergency telephone numbers, allergies, etc. They can also include a set of activities that are essential and/or enjoyable to the PLWD, such as personal hygiene; going for a walk; preparing lunch; gardening. Examples of the possible content of a care plan can be found in [18]. Should such activities be present in a care plan they can be included in the toolkit.

(iii) Enjoyable activities. The set of activities in this group will be tailored to each PLWD and are chosen because they bring pleasure to that person and/or have other benefits, such as the health benefits gained from an enjoyable walk, that can help maintain quality of life. Each PLWD, or their carer, will be asked to describe activities such as socialising, hobbies or exercising that the PLWD enjoys and wishes to maintain. From this list of activities, a set of activity goals will be agreed on and included in the individual's set of goals to be achieved. Additionally, a prepared set of activities, derived from our literature review and from co-design sessions attended by PLWD and their carers, will be presented to PLWD and their carers as prompts to ensure that they have not omitted any activities that they do enjoy but may not recall unprompted in interview. Two scales that measure pleasurable events or emotions in dementia were found: Teri and Logsdon's PES-AD [15], a set of activities and events created to measure the quality of life of individuals with Alzheimer's Disease, and Smith et al's DEMQOL [16]. Smith et al focused on recording changes in emotional states, whereas Teri & Logsdon focused on the accomplishment of actual events. Our list is based on the PES-AD scale. Example PES-AD activities include: Having friends visit; doing jigsaw puzzles; gardening; going to church.

ADL, IADL and PES-AD activity sets were each drawn up several decades ago and some activities included, for example making a telephone call, need to be updated to allow for communication using mobile phones, email, or social media. Other activities may need to take into account restrictions due to the current COVID-19 pandemic, e.g., 'going to church' may need to also allow for 'attending church services remotely'. Additional activities will be identified through co-design workshops attended by PLWDs and their family carers.

We aim to determine goals for each of the established ADLs and IADLs, and to measure each PLWD's achievement in reaching these goals. Further, for each PLWD, we take individualised goals from their care plans and from activities pleasurable to them, again measuring their achievement of these goals. Measurement of these goals will be either by PLWD or their family carer entering information directly into the toolkit interface, or by automated data collection by sensors.

For each activity included, for example 'ability to shop for oneself' or 'going to church', a decision needs to be made on the appropriate metric to determine achievement of each goal, and how to measure achievement of the goal. We aim to establish baseline target goal levels based on the PLWD's existing performance in order to maintain performance of activities and improve wellbeing.

Existing scales of daily activities and pleasurable activities can be utilised to create a set of goals for individuals with dementia. Achievement of these goals by PLWDs can benefit their quality of life and help them maintain their ability to live independently. Achievement of these goals can be directly reported by PLWDs or their family carers via a computerised toolkit's interface, or can be determined by data collected by sensors. Information recorded for each activity will include quantifiable information such as start time, end time, duration, achievement, comparison with goal target, and a measure of pleasure or satisfaction. For example, a PLWD may have had a daily walk starting at 10am and finishing at 10:45am, lasting 45 minutes, walking for 2.5 km and exceeding a target of 2 km. The measurable achievement score – in this case 45 minutes, 2 km – can be tracked over time, providing information to the PLWD and their carers about which areas of their daily life may require attention, potentially in the form of advice in improving achievement.

Examples of some goals are shown in Table 1. Included in the table is information on each goal, illustrating how we intend to quantify the achievement of each goal.

Each example in Table 1 shows a different type of goal. 'Preparing a meal', an IADL, is a basic activity we wish to encourage the PLWD to maintain as essential for continued independent living. 'Walking' is often included in care plans

TABLE 1. GOAL TARGETS AND THEIR MEASUREMENT

Activity source:	**(I)ADL/ MWDS**	**Care plan**	**PLWD/ Carer**
Specific activity:	Preparing a meal	Walking outside	Visiting a friend
Optionality:	Core	Core	Optional
Frequency goal:	Once a day	Once a day	Once a week
Duration goal:	20 minutes	60 minutes	120 minutes
Duration data source:	Sensors/ app	Sensor	App
Achievement goal:	Successful preparation	Walk 2.5 km	Visit friend
Average frequency:	5/week	6/week	0.8/week
Average duration:	24 minutes	55 minutes	140 minutes

 ISBN: 978-1-61208-916-4

for the health benefits it brings. 'Visiting a friend' is a little different from the previous two activities, as it is included as being chosen by the PLWD as a pleasurable activity that helps maintain their quality of life. Each activity has several dimensions that can be quantified: duration of activity, frequency of activity, activity achievement (e.g., distance walked, completeness of meal preparation). Measurement of these dimensions can be recorded by sensors, by asking the PLWD or their family carer to enter information on goal achievement into an app, or by a combination of the two. Using this approach allows for finer granularity of information to be captured than the traditional paper-based recording of ADL performance, and allows for the information to be captured in real-time or near-real-time rather than the usually retrospective information capture necessitated when information is gathered by visiting healthcare professionals.

IV. DISCUSSION

In order to record achievement of activities, whether (I)ADLs or pleasurable activities, data on achievement of activities must be captured. This can be done by sensors, with little or no demand on the PLWD, or by active data entry by the PLWD. We believe that of the 13 ADL and IADL activities, 8 can be monitored by use of sensors (e.g. 'bathing'), leaving 5 that will require manual entry by the PLWD (e.g. 'shopping'). The intention of the co-design methodology is that, by engaging with PLWDs at the app design stage, the app will be simple to use and be seen by the PLWD as a device of empowerment rather than a burden. By including activity reminders and goal targets it is intended that the app be a useful tool, and our early feedback from carers is that this will be the case. However, we await the outcome of the full co-design sessions and feedback on use of the developed app to confirm or otherwise that this is the case.

There is some literature on behaviour change, such as the work of Fogg et al [19], Oinas-Kukkonon [20], Reimer et al [21] and Webb et al [22], which will inform our work on assisting PLWD to continue to maintain good behaviours. in particular, Fogg et al's categorisation of new behaviours into 'one-time', 'temporary' or 'permanent'. We take further inspiration from the transtheoretical model of health behavior change of Prochaska and Velicer [25], in particular because we are aiming at maintaining activities of PLWDs rather than changing them.

Success of our system will be measured by recording period of independent living against existing times, rate of decline in performing activities, and adherence to self-reporting of activity performance. Limitations on our work are the engagement of PLWDs and their carers, the technology challenges of tracking activities by sensors, and the appropriate selection of activities and activity achievements.

V. CONCLUSIONS

A set of goals for PLWDs can be constructed from existing scales of performance and disability, including scales that are not specific for dementia and those that are, and from care plans. Additional goals can be included following discussion at co-design workshops and interviews with individual PLWD. Goal targets can be determined from care plans, from existing achievement levels of PLWD and achievement ambitions. The set of goals and goal targets can be used to construct a framework that will allow for computational models to determine goal achievement and suggestions to maintain or achieve goal targets.

Future work will include developing methods of capturing goal achievement information with minimum burden to the PLWD, by using an app with a simple interface and/or capturing and processing data automatically from sensors; developing appropriate methods of prompting or encouraging PLWDs to achieve or repeat goals, and mapping relationships and dependencies between goals (for example, the ability to walk or drive a minimum distance, or the ability to use public transport, which may be necessary for the ability to perform independent shopping). A pilot implementation of the system in the homes of 6 PLWDs is planned, to include evaluation of activity achievement, and burden on the PLWDs and their carers.

ACKNOWLEDGMENT

This material is based upon works supported by the Science Foundation Ireland under Grant No. 19/FFP/6917.

REFERENCES

[1] Health Service Executive. *Types of Dementia*. [Online]. Available from: https://www.understandtogether.ie/about-dementia/what-is-dementia/types-of-dementia/ [retrieved: September, 2021]

[2] M. Prince, A. Comas-Herrera, M. Knapp, M. Guerchet, and M. Karagiannidou, *World Alzheimer report 2016: improving healthcare for people living with dementia: coverage, quality and costs now and in the future*. [Online]. Available from: https://www.alzint.org/u/WorldAlzheimerReport2016.pdf [retrieved: September, 2021]

[3] M. Prince et al, *World Alzheimer Report 2015 - The Global Impact of Dementia: An analysis of prevalence, incidence, cost and trends.* [Online]. Available from: https://www.alzint.org/u/WorldAlzheimerReport2015.pdf [retrieved: September, 2021]

[4] C. Olsen et al, "Differences in quality of life in home-dwelling persons and nursing home residents with dementia - a cross-sectional study," BMC Geriatr. 2016 Jul 11;16:137.

[5] Department of Health. *The Irish National Dementia Strategy . Department of Health: Dublin.* [Online]. Available from: https://www.hse.ie/eng/about/who/healthwellbeing/healthy-ireland/publications/irish-dementia-strategy-1-.pdf [retrieved: September, 2021]

[6] S. Katz, A. B. Ford, R. W. Moskowitz, B. A. Jackson, and M. W. Jaffe, "Studies of illness in the aged. The Index of ADL: a standardized measure of biological and psychosocial function," JAMA 185, pp:914-919, 1963

 ISBN: 978-1-61208-916-4

[7] M. P. Lawton and E. M. Brody, “Assessment of older people: Self-maintaining and instrumental activities of daily living,” The Gerontologist, 1969 9(3), pp:179-186.

[8] NHS South East Clinical Delivery and Networks. *Dementia Care and Support Planning Generic Toolkit.* [Online]. Available from: https://www.southeastclinicalnetworks.nhs.uk/dementia-csplanning-toolkit/ [retrieved: September, 2021]

[9] L. Clare et al, “Individual goal-oriented cognitive rehabilitation to improve everyday functioning for people with early-stage dementia: A multicentre randomised controlled trial (the GREAT trial),” Int J Geriatr Psychiatry. 2019 May;34(5), pp: 709-721.

[10] L. Clare, “Rehabilitation for people living with dementia: a practical framework of positive support,” PLoS Med. 2017;14(3):e1002245.

[11] L. Clare et al, “Goal-oriented cognitive rehabilitation in early-stage dementia: study protocol for a multi-centre single-blind randomised controlled trial (GREAT),” Trials. 2013 May 27;14, pp: 152-165.

[12] A. Phinney, H. Chaudhury, and D. L. O'Connor, “Doing as much as I can do: the meaning of activity for people with dementia,” Aging Ment Health. 2007 Jul;11(4), pp: 384-93.

[13] T. J. Kiresuk and R. E. Sherman, “Goal attainment scaling: A general method for evaluating comprehensive community mental health programs,” Community Ment Health J. 1968 Dec;4(6), pp: 443-53.

[14] F. Mahoney and D. Barthel, “Functional evaluation: the Barthel Index,” Md Med J. 1965;14, pp: 61–65.

[15] L. Teri, and R. G. Logsdon, “Identifying pleasant activities for Alzheimer's disease patients: the pleasant events schedule-AD,” Gerontologist. 1991 Feb;31(1), pp:124-7.

[16] S. C. Smith et al, “Measurement of health-related quality of life for people with dementia: development of a new instrument (DEMQOL) and an evaluation of current methodology,” Health Technol Assess 2005;9(10).

[17] I. P. Donald, "Development of a modified Winchester disability scale--the elderly at risk rating scale," J Epidemiol Community Health, 1997;51(5), pp:558-563.

[18] Alzheimer’s Association. *Daily Care Plan*. [Online]. Available from: https://www.alz.org/help-support/caregiving/daily-care/daily-care-plan [retrieved: September, 2021]

[19] B. Fogg and J. Hreha, “Behavior wizard: A method for matching target behaviors with solutions,” in: International Conference on Persuasive Technology, Springer, 2010, pp. 117–131.

[20] H. Oinas-Kukkonen, “Behavior change support systems: The next frontier for Web science,” In: Proceedings of the Second International Web Science Conference (WebSci 10), Raleigh, NC, US, April 26-27 2010.

[21] U. Reimer, E. Maier and T. Ulmer, “A Self-Learning Application Framework for Behavioral Change Support,” In: C. Röcker, J. O’Donoghue, M. Ziefle, M. Helfert, W. Molloy (Eds.): Information and Communication Technologies for Ageing Well and e-Health. Second International Conference, ICT4AWE 2016, Revised Selected Papers. Springer, 2017, pp. 119-139.

[22] T. L. Webb, J. Joseph, L. Yardley and S. Michie, “Using the internet to promote health behavior change: A systematic review and meta-analysis of the im- pact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy,” Journal of Medical Internet Research 12 (1) (2010) e4.

[23] J. O. Prochaska, and W. F. Velicer, "The transtheoretical model of health behavior change," American Journal of Health Promotion, 1997;12(1), pp: 38–48.

 ISBN: 978-1-61208-916-4

First Iteration Test of the Navigation User Interface from ADAPEI Transport App with Adults Having Intellectual Disabibilities

Jesus Zegarra Flores
Research and innovation Department
Capgemini Engineering
Illkirch-Graffenstaden, France
e-mail: jesus.zegarraflores@altran.com

Emma Charbonnier
Research and innovation Department
Capgemini Engineering
Illkirch-Graffenstaden, France
e-mail: emma.charbonnier@capgemini.com

Nadia Laayssel
Centre Ressources Enfance & Adolescence
ADAPEI du Territoire de Belfort
Belfort, France
e-mail: n.laayssel@adapei90.fr

Rémi Coutant
Centre Ressources Enfance & Adolescence
ADAPEI du Territoire de Belfort
Belfort, France
e-mail: r.coutant@adapei90.fr

***Abstract*— People with intellectual disabilities face cognitive problems that affect their memory, spatial and time perception. The "ADAPEI Transport" app has been created by Capgemini engineering and the ADAPEI association from Belfort in France. It is a tool to help children and young adults having intellectual disabilities to learn how to use a public transport and to walk in their own cities independently. The reference route is first physically (in situ) created by a specialist. Along the path, steps are recorded to help the user with intellectual disabilities to interact with his/her environment in the navigation part. The aim of this article is to show the results of a first iteration feedback done with three adults having intellectual disabilities and a specialist in mobility to improve and adapt the app to adult's special needs to walk and to take the public transport independently in unknown environments.**

Keywords-GPS; intellectual disability; user interface, mobile app.

I. INTRODUCTION

The ADAPEI (association of friends and parents of mentally disabled people in France) is working with educators to teach intellectual disabled children (aged 10 to 20) to become more autonomous to take public transport. Specialists are currently working with paper leaflets to teach them which point of interests to look for and what action to do (pedestrian crossing, stop a bus, walk to a bus station, etc.) for the journey. The creation of leaflets and learning process can be long, a solution to accelerate this learning has been the creation of the ADAPEI Transport app [1].

Currently, most navigation tools are not accessible enough, to people having intellectual disabilities. Most of the mobile apps rely on Google Direction API to define simple instructions to guide the user. An application is WaytoB, which serves to guide the disabled person with his/her phone and a smartwatch [2], the person is tracked on real time by his/her caregiver thanks to internet connection. Another mobile app is AssisT-OUT [3], which provides street views from Google Maps to help the user to recognize his/her environment and to take decisions. Indeed, these mobile apps do not focus enough on the learning notion, which is essential to help children to become independent. In this article, we will show a first iteration test done with adults having intellectual disabilities for improving and adapting the current ADAPEI transport app developed for children to adults who work and are autonomous in known paths and environments.

In Section 2, authors will explain the current ADAPEI transport app, the different user interfaces and how the information is delivered to have feedback form end-users. In Section 3, feedback results from 3 adults having intellectual disabilities and a specialist are presented. Finally, the conclusions and the future work are presented in Section 4.

II. METHODOLOGY

In this section, it is going to be presented a brief explanation of the current adapei transport app, some information about the user navigation interface and the kinds of pedagogical videos created to be shown to adults having intellectual disabilities.

A. ADAPEI Trasnport app

The app (created in Android) has been developed in narrow collaboration with the specialists and young adults from the transport workshop from the ADAPEI's Belfort local branch. It has been tested constantly from 2018 with end users (aged from 10 to 20) to meet their special needs. The app allows the specialist to create an adapted reference path using GPS, pictograms, photos, times and voice message indicating the action to do. The app also has a navigation part in which a specific user interface has been developed delivering information of the whole path selected and from every step of this path when the person is close to a landmark.

B. The navigation user interface

Firstly, the app recovers the reference path selected already recorded by the specialist. The user interface of the navigation part is presented in a way of list showing all the steps to do (Figure 1) with pictograms, photos and vocal message. The GPS starts working and when the person is close to the zone of a landmark step (less than 30 meters), the

 ISBN: 978-1-61208-916-4

smartphone will vibrate and will show in a bigger size the current step information to do on the screen (for example, crossing the street or going to the bus station Emile Mathis, get into the bus at 9 o' clock, etc.) compared to the other steps.

Additionally, it will launch the voice message information if it is demanded. When the person has finished the action, the current action disappears and the next action to do will appear in a bigger size. Many algorithms and GPS data filters have been implemented to understand when the person has finished an action and that a new action has to be done. A video of the app can be watched in [4].

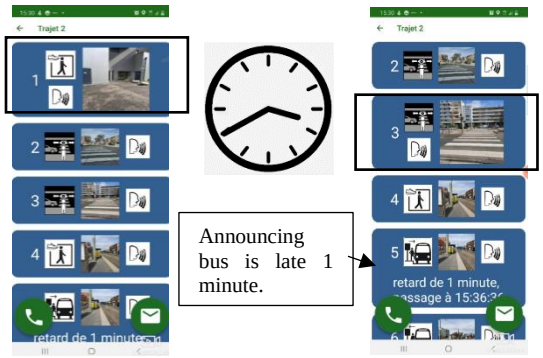


Figure 1. Navigation user interface in a list mode.

In the Figure 1 above we can see a sequence, once the person has done the first action, after walking in the path, the closest current action is crossing the street (3) but before he/she has to cross a previous street (2) (steps 2 and 3 are very close).

C. *Creation of the pedagogical videos*

The video was created and edited in order to show in a pedagogical way how the information appear during navigation. Two paths were created as reference paths using the app recording all the information. After, a navigation simulation is created using the navigation interface screen synchronized with the video from the environment showing the information presented to walk from one-step to another. We have edited two videos: one having the information of photos of landmark, pictograms and the actions to do (see Figure 2) and another who has photos, pictograms and arrows to indicate to turning right or left in an intersection.

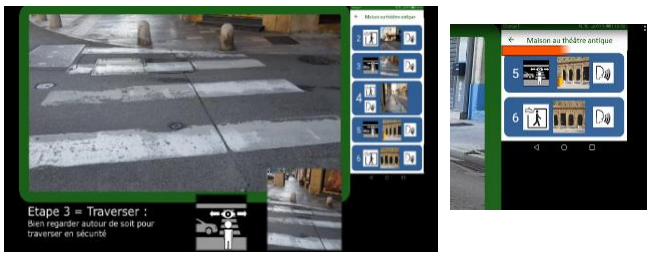


Figure 2. On the left, crossing the street (step 3) from a video, on the right progression bar (the person has walked half of the distance going to the step 5)

Because of the COVID sanitary problem, these videos were presented using Microsoft Teams meetings.

In the focus group, there are three adults having intellectual disabilities who are able to walk independently in known paths and environments and a specialist in mobility. The 3 disabled subjects are older than 20 years old. One of them, is not familiar with the use of smartphones.

The paths shown in the videos and the app are unknown for the subjects. (an example of the video shown can be found in [5]).

In the TEAMS meeting, researchers have explained the different information (pictograms and photos) to them and we have asked to the participants what they will do in every step. We have taken notes from their responses.

III. RESULTS

In this part, we are going to present the feedbacks from the adults and a specialist.

The progression bar showing the distance walked by the person was not very visible. It should be bigger and with more contrast color compared to the background.

All of them agree that images and pictograms are more understandable than turning right and turning left arrows on the screen.

Currently, the landmark steps of the path disappears automatically once the person has finished the action and gets out from the step zone. For the specialist, one improvement is not to make the image of the step disappears, in contrast, to validate a landmark step it could be more convenient to put on the step a big X cross once the step has been done by the user.

Currently, the validation when a person is in the zone of a step is done 30 meters before the GPS coordinates recorded of the step position. Users ask if that could be done the closest possible to the step position to avoid confusion when 2 steps are very close.

A user who already uses Google maps in cars has told us that one of the problems about car navigation is that they do not have the photo or image of the destination. An advantage of our app is that we can create and edit a specific path with the photo of the landmarks and the destination.

According to the specialist, the subjects seem not to have problems of comprehension of the environment but they could experience lack of self-confidence in new environments. The app can help them to be reassured.

All the subjects would also like to be informed that they do not take the right direction in the path (for example, in case of intersection).

IV. CONCLUSIONS AND FUTURE WORK

In this article we have presented a brief explanation of the adapei transport app and the pedagogical videos edited to show the possibilities to adapt the user interface app to the special needs of adults having intellectual disabilities.

In general, the three people having intellectual disabilities seem to have understood, after asking them, the indications in our video and all the steps using the images and the pictograms.

The app can be used to reassure the person that he/she has taken the right path especially in unknown environments.

 ISBN: 978-1-61208-916-4

For giving information about if he/she has taken the wrong path in an intersection, the use of the information of the compass of the phone could be integrated.

After this first iteration, we will rework on the code to add and change the parts in the app to improve the user interface.

The next iteration will consist on testing the user interface of the app with end-users using the path creation and navigation user interfaces in a real situation.

REFERENCES

[1] J. Zegarra Flores, G. Malnati, J.J. Stevens, E. Bournez, L. Boutros, N. Laayssel, G. Geneviève, J.B. de Vaucresson, R. Coutant, J.P. Granger and J.P. Radoux. "ADAPEI-TRANSPORT: A GPS Based Mobile App for Learning Paths and Improving Autonomy for Young Adults Having Intellectual Disabilities to Take Public Transport", Computers Helping People with Special Needs. ICCHP 2020. Lecture Notes in Computer pp. 112-119.

[2] R. Fryers, T. Holzer Saad, K. Kelly and J. Dinsmore. "Report Defining The Needs of Stakeholders For a Wayfinding Platform For Individuals With Intellectual Disabilities and Their Carers", Trinity College Dublin, Trinity College Dublin, 2018, pp. 1-29.

[3] J. Gomez, G. Montoro, J. C. Torrado, and A. Plaza. "An Adapted Wayfinding System for Pedestrians with Cognitive Disabilities" in Mobile Information Systems, Volume 2015, Article ID 520572, pp. 1-11.

[4] https://www.youtube.com/watch?v=hJp-qIrbNTY

[5] https://www.youtube.com/watch?v=5-ZHv833jdk

 ISBN: 978-1-61208-916-4

Towards the Support of Design Patterns in the Fast Healthcare Interoperability Resources (FHIR) Standard

Timoteus Ziminski
Department of Computer Science and Engineering
University of Connecticut
Storrs, USA
e-mail: timoteus.ziminski@uconn.edu

Steven Demurjian
Department of Computer Science and Engineering
University of Connecticut
Storrs, USA
e-mail: steven.demurjian@uconn.edu

Thomas Agresta
Department of Family Medicine
University of Connecticut Health Center
Farmington, USA
e-mail: thomas.agresta@uconn.edu

Edward VanBaak
University College
University of Denver
Denver, USA
e-mail: edward.vanbaak@du.edu

***Abstract*— The Fast Healthcare Interoperability Resources (FHIR) advanced the Health Level 7 v2 (HL7), Extensible Markup Language v3 (XML), and Clinical Document Architecture (CDA) standards by defining over 135 resources that conceptualize the different aspects of healthcare data to facilitate the secure exchange of data for patients using cloud-based Application Programming Interfaces (APIs). As part of the design and development process of FHIR applications, a subset of the resources would be chosen to satisfy application requirements. While the FHIR standard is excellent at conceptualizing resources, this process is very targeted at the implementation level. FHIR provides no higher-level constructs to organize resources in a predevelopment process similar to what is available in software design patterns, which arose when developers noticed that they were constantly reimplementing the same types of functionalities in the same way for different applications. Our prior work extended the FHIR standard with meta resources, a conceptual construct that defines the involved resources and their interactions into one unified artifact akin to a design pattern, with the ability to generate a FHIR XML schema bundle that could serve as the foundation in a development workflow. This paper examines the way that our previously defined meta resource extension to FHIR can be positioned within the FHIR standard by: considering the relation between meta resources and resources that extends the FHIR Unified Modeling Language (UML) model; assessing the ability of the abstract DomainResource to implement containment in meta resources; exploring FHIR modules that provide organizational and contextual information different resources; examining the way that meta resources fit into the FHIR information architecture; and, determining if it is possible to employ a composition of Library resources and ActivityDefiniton resources in support of realizing meta resources.**

Keywords—design patterns; FHIR standard; interoperability; data design

I. INTRODUCTION

The Fast Healthcare Interoperability Resources (FHIR) [1] was introduced in 2014 to support and continue to advance interoperability and Health Information Exchange (HIE) and promote the secure sharing of healthcare data. The standard has passed through an initial maturity level and is currently widely endorsed and accepted by stakeholders, policymakers, and healthcare application developers. As a prominent example, The Office of the National Coordinator mandates FHIR [2] for making data available via APIs to patients and for data sharing. The FHIR standard defines over 135 resources, which are data elements to capture all types of healthcare data organized into different layers; the *base resource* layer contains patients, practitioners, family relationships, organizations, services, appointments, and encounters; and, the *clinical* layer has resources for a patient's health history, diagnostic data, medications, care provision, and request/response communication. Resources are also organized in the *foundation*, *financial*, and *specialized* layers.

In support of HIE and interoperability, we have prior work [3] on software architectural alternatives (e.g., service-oriented, grid computing, publish & subscribe paradigm, and data warehousing) with follow-up work [4] using FHIR to integrate health information technology (HIT) systems to facilitate collaboration among medical providers. Our most recent work [5] explored the inclusion of a design pattern-like concept of a *meta resource* in the FHIR standard. Meta resources can logically organize FHIR resources into a conceptual unit that, like a design pattern, can be customized and reused across multiple applications. With a collection of meta resources, it becomes possible to take design decisions in application development on a higher conceptual level, lower the overall development effort through the reuse of knowledge and introduce tool support and automated requirement validation of implemented solutions. Our focus in this paper is to explore integrating the meta resource concepts directly into the FHIR standard.

Design patterns are a classic and widespread approach in software engineering and development pioneered in [6]. The idea originated in an investigatory process when software engineers and developers realized that they constantly copied and reimplemented similar code in different projects. A well-known example is the Observer pattern for defining a one-to-many relationship between a *subject* (active object) and *observers* (passive objects), in which observers are notified about changes to the subject. The Observer pattern is instantiated to solve the problem of having one object change multiple other objects (e.g., reacting on an incoming HTTP request). Another widely used example is the Model View Controller pattern, which consists of: a *model* that captures the Enterprise data and the business rules associated with accessing and updating the data; a *view* that renders the contents or subset of the model for the presentation of the stored data; and, the *controller* that translates interactions with the view (button clicks, user interface (UI) actions, Hypertext Transfer Protocol (HTTP) calls, etc.) into actions on the model. This initial work subsequently led to the definition of higher-level architectural design patterns [7] that are more conducive to use application development using FHIR to facilitate HIE.

 ISBN: 978-1-61208-916-4

Design patterns can fit into the Enterprise architecture [8] of FHIR [9] as part of its overall information architecture. In addition to resources organized in the aforementioned five layers, the broader FHIR architecture model also includes a sixth *resource contextualization* layer containing FHIR profiles and graphs. Meta resources can fit into the FHIR's resource contextualization layer, thereby extending its architecture with additional relationship types, attributes, and constraints. Our recent work [5] leverages the extension concepts of FHIR profiles along with the grouping operation of Bundle resources to integrate meta resources seamlessly into the FHIR standard. This approach provides a design pattern-like capability that can augment and extend the existing functionality of FHIR by higher-level conceptual named constructs that, for particular workflows, clearly define the involved resources and their interactions into one unified artifact.

The main objective of this paper is to provide a detailed examination of the way that our meta resource extension to FHIR [5] can be positioned within five facets of the FHIR standard. The first facet is the inclusion of meta resources that would extend the FHIR UML model of resources [10]. The second facet will assess whether the abstract DomainResource [11] can be leveraged to implement containment in meta resources. The third facet explores FHIR modules [12] to see if they can provide the organizational and contextual information required for meta resources. The fourth facet examines the way that meta resources could fit into the FHIR information architecture [9]. The final facet explores the extent to which meta resources can be realized as a composition of Library resources [13] and ActivityDefiniton resources [14].

To illustrate the potential of the concepts of this paper, we will focus our work on the development of a mobile health application and framework [15] for medication reconciliation that integrates information from multiple EHRs. Please note that FHIR is an evolving standard (normative version v.4.0.1 at the time of writing). Therefore, the concepts in this paper and a discussion of the facets may be subject to change in future releases. As a result, the paper explores these facets as alternative approaches for incorporating meta resources into the FHIR standard instead of suggesting one fixed approach.

The remainder of this paper has four sections. Section 2 provides background information on medication reconciliation concepts and the FHIR concepts, both of which are used for examples throughout the paper. Section 3 summarizes our prior work [5] to present an outline of our model that extends FHIR with meta resources from which FHIR bundles can be automatically generated. Section 4 explores the aforementioned five facets to fully understand the way that meta resources can be included in the FHIR standard, thereby supporting a design pattern-like capability. Section 5 contains concluding remarks and outlines ongoing work.

II. BACKGROUND

This section provides background material in two areas. Section II.A explores medication reconciliation and its importance for healthcare. Section II.B presents FHIR concepts relevant to the paper.

A. Medication Reconciliation

Since a patient's medication regimen is the basis for treatment decisions, medication lists must be accurate to maximize therapeutic impact and prevent potentially life-threatening events. Discrepancies between the medication lists in HITs where patients receive care can potentially harm patients. This challenge is significant enough that in Connecticut (CT), the CT General Assembly passed Special Act 18-6: An Act Requiring the HIT Officer to Establish a Work Group to Evaluate Issues Concerning Polypharmacy and Medication Reconciliation [16] which produced several recommendations to the legislature which includes the development of technology to support the ability to generate the Best Possible Medication History via an automated electronic means [17]. To substantiate the discussion in this paper, we provide examples from our work [15], which presents a medication reconciliation solution that integrates information from multiple EHRs. We can combine the FHIR resources used by this solution (Patient, Medication, MedicationStatement, etc.) into a MedicationReconciliation meta resource which supports the actions of the application, including: retrieving medications from multiple EHRs, personal health records, and other HITs; combining and reconciling medications into a best medication list that identifies potential conflicts between the same or different medications; and, supporting an adaptive multi-use algorithm for medication reconciliation.

B. The Fast Healthcare Interoperability Resource, FHIR

FHIR is structured around the concept of resources, which are comprehensive data elements that hold the information expressed in FHIR. Formally, a FHIR resource R is defined as an entity with the properties set P = (Identity, Type, (Data Item*), Version). The *identity* property is used to address a given resource entity within a FHIR system consisting of one or more FHIR servers. The *type* property specifies one of the resource types that are provided by the FHIR specification. The *data items* property is a set of structured data elements, which holds the resource's actual data content as specified by its definition. The *version* property implements a version counter which tracks changes that occur to the contents of a resource through its lifetime. The record version automatically changes each time the resource changes, allowing a complete audit trail that tracks the evolution and the provenance of a resource. The FHIR standard defines representation formats in XML, JavaScript Object Notation (JSON), and Terse RDF Triple Language (Turtle).

```
<Medication xmlns="http://hl7.org/fhir">
 <identifier><!-- 0..* Identifier -->
 </identifier>
 <code><!-- 0..1 CodeableConcept --></code>
 <status value="[code]"/><!-- 0..1 -->
 <manufacturer>
  <!-- 0..1 Reference(Organization) -->
 </manufacturer>
 <form><!-- 0..1 CodeableConcept --></form>
 <amount><!-- 0..1 Ratio --></amount>
 <ingredient>
  <!-- 0..* Active/inactive ingredient -->
  <item[x]>
  <!-- 1..1 CodeableConcept|
   Reference(Substance|Medication) -->
  </item[x]>
  <isActive value="[boolean]"/><!-- 0..1 -->
```

 ISBN: 978-1-61208-916-4

```
  <strength><!-- 0..1 Ratio --></strength>
 </ingredient>
 <batch>
  <!-- 0..1 Packaged medications details -->
  <lotNumber value="[string]"/><!-- 0..1 -->
  <expirationDate value of="[dateTime]"/>
   <!-- 0..1 -->
 </batch>
</Medication>
```

Figure 1: Medication XML Schema.

Figure 1 contains an abbreviated portion of the Medication resource. Note that for these two examples and all other examples, we have omitted some of the details as it impacts the single column display; see [18] for complete versions.

III. SUMMARIZING META RESOURCE CONCEPTS

This section reviews the meta resource model of our prior work [5], highlighting the key concepts necessary for the remainder of the paper. Section III.A reviews the model for describing a meta resource and provides examples. Section III.B explores the generation of a FHIR bundle that, combined with the meta resource definition, represents a design pattern-like for a particular problem.

A. Meta Resource Capabilities and Example

To begin, a *FHIR resource* is a four-tuple of identifier, type, date, and version $R = < R_{ID}, R_{Type}, R_{Data}, R_{Version} >$. Example 1 illustrates a Patient [18] instance according to this definition.

Example 1. A FHIR *Patient* resource instance for patient John Doe after five changes is represented by

$$R_1 = < R_{ID_1}, t_1, R_{Data_1}, x >$$

where

$$R_{ID_1} = http://test.fhir.org/rest/Patient/123$$

and

$$t_1 = Patient$$

and

$$R_{Data_1} = \{\{"identifier" : "ea44426f",$$
"active": "true",
"name": "John Doe",
"telecom": "555-370-8047",
"gender": "male",
$"birthDate" : "1970 - 12 - 12", \ldots\}$

and

$$x = 5$$

Next, we review the concept of *meta resources* to leverage the design pattern idea and define higher-level design constructs that can represent multiple resources needed to implement a particular application. This transcends the resource-centric view of FHIR on clinical data. Meta resources provide reusable workflow-centric patterns that allow a conceptual view for implementing functionalities in an already FHIR enabled system. Sample workflows include medication reconciliation, patient admission, or vaccination support.

The definition of a specific meta resource determines the properties, components, relationships, and requirements that the given meta resource has to an implementing system. To implement a particular workflow, a meta resource from a library of previously implemented solutions can be used as a pattern for the high-level design of applications and as a schema against which actual implementations are checked for full functionality. The overall objective of a meta resource is the larger granular organization of resources for a specific health-related workflow, described via human-readable description.

Example 2 has an instance of a *MedicationReconciliation* meta resource for the medication reconciliation as introduced in Section II.A and implementing the meta resource structure shown in Figure 2: the Patient resource holds demographic information of the affected patient. It references one or more MedicationStatement resources. The MedicationStatement identifies that a patient is or was taking a medication. It contains a Medication resource identifying the actual medication. It also references one Endpoint resource to indicate from where the statement was retrieved. Finally, it also references one Practitioner resource that identifies the individual who should be notified if issues regarding this statement are detected. A DetectedIssue resource references two or more medication statements, indicating a potential problem between those statements, which must be resolved during reconciliation. An Endpoint records information on which system needs to be notified regarding a discovered issue. It represents a health information technology system such as an electronic health record at a hospital, a rehab facility, a clinician's office, etc., that must be notified whenever a change is made. The Endpoint references a Practitioner resource to identify a practitioner who needs to be contacted regarding issues detected during reconciliation for a given medication statement. In summary, the MedicationReconciliation meta resource comprises six different FHIR resource types and models their relationships in the scope of a medication reconciliation workflow.

Example 2. In the formal model introduced in [5], the meta resource *MedicationReconciliation* is represented by

$$MR_1 = < MR_{ID_1}, MR_{Name}, MR_{Desc_1}, MR_{PR_1}, MR_{API}, MR_{REF_1}, MR_{COM_1} >$$

where

$$MR_{ID_1} = http://test.fhir.org/rest/meta/MedicationReconconciliation/123$$

and

$$MR_{Name_1} = MedicationReconciliation$$

and

MR_{Desc_1} = "*Medication reconciliation is the process of comparing a patient's medication orders to all the medications that the patient has been taking and eliminating potential errors, resulting in a new up to date list of medications.*"

and

$$MR_{PR_1} = \{\text{pr}_{\text{patient}} = < Patient, \text{entity} >),$$
$$\text{pr}_{\text{medStatement}} = < MedicationStatement, \text{producer} >,$$
$$\text{pr}_{\text{med}} = < Medication, \text{dataSource} >,$$
$$\text{pr}_{\text{endPoint}} = < EndPoint, \text{address} >,$$
$$\text{pr}_{\text{practiotioner}} = < Practitioner, \text{entity} >,$$
$$\text{pr}_{\text{issue}} = < DetectedIssue, entity >\}$$

and

$$MR_{API} = < profile >$$

and

$$MR_{REF_1} = \{< pr_{patient}, pr_{medStatement} >, < pr_{medStatement}, pr_{endPoint} >, < pr_{medStatement}, pr_{practitioner} >, < pr_{issue}, pr_{medStatement} >\}$$

 ISBN: 978-1-61208-916-4

and

$$MR_{COM_1} =< pr_{medStatement}, pr_{med} >$$

Each resource in a meta resource is classified according to the service it provides, such as *consumer*, *producer*, *data source*, or *data generator*. For example, in Figure 2, the Medication resource will serve as a *data source* for performing the reconciliation. In contrast, the FHIR *MedicationStatement* can serve as a *producer* of further medication resources. Fundamentally, a meta-resource is a composition of standard FHIR resources enriched with meta-information to define the structure and interactions of FHIR resources on the design level. A meta-resource definition contains a specification of FHIR resources that must be available to instantiate the meta resource, known as the participating resources.

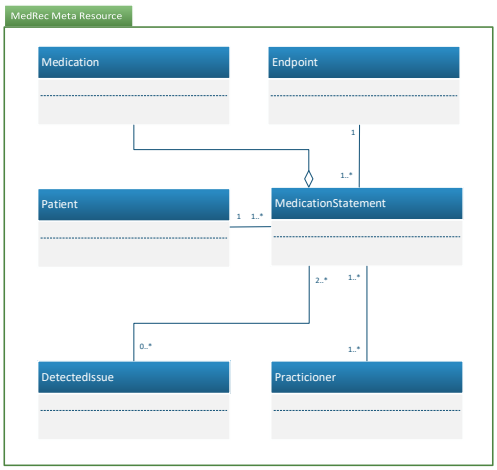


Figure 2. Medication Reconciliation Meta Resource.

B. FHIR Bundle Generation

This section briefly reviews our approach for automatically transitioning a meta resource definition into a destination FHIR bundle at the schema level. Details of the process and the associated algorithm were documented in our previous work [5]. FHIR Bundle resources [19] are container resources built explicitly into the standard to group other references, for example, in the context of search results or for the exchange of messages. Conceptually, a bundle consists of meta-information about the bundle itself (such as an identifier), entries encapsulating FHIR resources, and links between those entries and potentially other outside resources. FHIR bundles allow references and containment and are capable of reflecting the relationships shown in the meta resource in Figure 2.

As one transitions from design to the development of the healthcare application, the FHIR bundle can facilitate the exchange with another system. This approach aims to ensure baseline compatibility with systems that are not aware of meta resources while simultaneously leveraging existing communication APIs for interacting with other meta resource-enabled systems. To illustrate the process, a meta resource needs to be transitioned to a suitable FHIR artifact to construct a representation that is amenable to realization within a particular healthcare application.

```
<Bundle xmlns="http://hl7.org/fhir">
 <!-- for readability expand $url = http://hie.org/fhir -->
 <id value="bundle-metaressource-collection"/>
 <type value="collection"/>
 <entry>
  <fullUrl value="$url/Patient/123"/>
  <resource><Patient><!-- John Doe --></Patient></resource>
 </entry>
 <entry>
  <fullUrl value="$url/Practioner/123"/>
  <resource><Practitioner><!-- Dr. Max Mustermann --></Practitioner></resource>
 </entry>
 <!-- Further practioner entries -->
 <entry>
  <fullUrl value="$url/Endpoint/123"/>
  <resource><Endpoint><!-- hl7-fhir-rest endpoint --></Endpoint></resource>
 </entry>
 <!-- Further endpoint entries -->
 <entry>
  <fullUrl value="$url/MedicationStatement/123"/>
  <link><relation value="related"/><url value="$url/Patient/123"/></link>
  <link><relation value="related"/><url value="$url/Practioner/123"/></link>
  <link><relation value="related"/><url value="$url/Endpoint/123"/></link>
  <resource>
   <MedicationStatement>
    <id value="medstatement001"/>
    <contained><Medication> <!-- id="med309" --></Medication></contained>
    <medicationReference><reference value="#med0309"/></medicationReference>
    <subject><reference value="Patient/123"/></subject>
   </MedicationStatement>
  </resource>
 </entry>
 <!-- Further MedicationStatement entries, including
 medstatement002 at $url/MedicationStatement/124 -->
 <entry>
  <fullUrl value="$url/DetectedIssue/123"/>
  <link><relation value="related"/><url value="$url/MedicationStatement/123"/></link>
  <link><relation value="related"/><url value="$url/MedicationStatement/124"/></link>
  <resource>
   <DetectedIssue>
    <code><!-- Drug Interaction Alert --></code>
    <implicated><reference value="MedicationStatement/medstatement001"/></implicated>
    <implicated><reference value="MedicationStatement/medstatement002"/></implicated>
   </DetectedIssue>
  </resource>
 </entry>
 <!-- Further DetectedIssue statements -->
</Bundle>
```

Figure 3. Generated Bundle.

Our algorithmic process documented in [5] takes the meta resource of Figure 2 and creates a meta resource bundle that captures all of the components and relationships conforms to the FHIR bundle XML schema and generates a bundle resource as given in Figure 3. Specifically, for every medication reconciliation meta resource, the initial generation is the top-level bundle element and a single patient element. Then the participating resources, as given in Figure 2, are processed into matching bundle entries and explicitly connected by bundle links to express their interactions and relationships.

IV. INTEGRATING DESIGN PATTERNS INTO FHIR

This section reviews five facets of the FHIR standard that give anchor points for extending the standard and explores multiple approaches for conceptually including the meta resource model concepts of Section III.A into the standard. The first facet in Section IV.A investigates the relation between meta resources and FHIR resources, which is explored by extending the relevant UML model. Section IV.B, the second facet, inspects the functionality provided by the abstract DomainResource resource as a potential candidate to implement the containment of resources in meta resources. Section IV.C, the third facet, explores the organization of FHIR resources in FHIR modules and identifies modules that would be affected by the meta resource extension. Section IV.D, the fourth facet, details the way that meta recourses fit into the FHIR information architecture. The fifth facet in Section VI.E explores the composition of the Library and ActivityDefiniton resources. The concepts discussed in this section are partially at the initial stages of development (e.g., the FHIR module definitions) and may be subject to change in future releases. Therefore, we explore several alternative approaches for incorporating meta resources instead of suggesting a fixed direction.

 ISBN: 978-1-61208-916-4

A. Meta Resource Capabilities and Example

To begin, in the first facet we consider, meta resources are directly related to the published resources [10] defined in the FHIR standard. They assemble a set of participating resources enriched with additional context information into higher-level artifacts. On a schema level, this enhances reusability by tailoring extensive resources that can be used to solve specific workflows in different parts of the health care domain. This is similar to the way that design patterns are used to tackle reoccurring problems across the software engineering domain. Then on an instance level, meta resource instances directly group instances of FHIR resources containing medical data and share them across systems that use FHIR seamlessly. This direct relationship between meta resources and participating FHIR resources allows for incorporating meta resources into FHIR by regarding them as entities related to FHIR resources. Figure 4 shows the UML structure of meta resources in the context of the FHIR resource UML structure, with new UML classes added for MetaResource and ResourceTuple at the top of the figure (top two boxes).

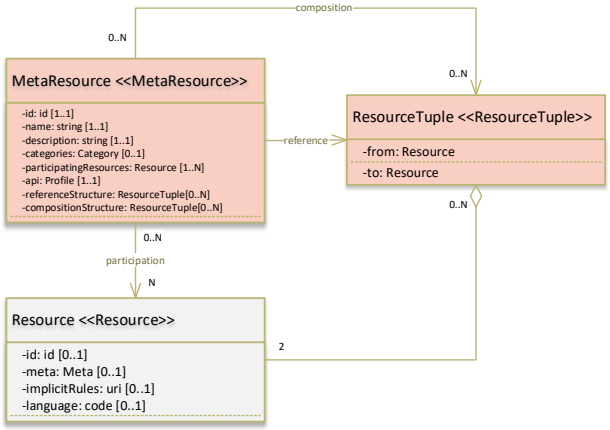


Figure 4. Meta Resource UML Structure.

Figure 4 demonstrates the UML artifact of a FHIR resource class defined by the FHIR specification [10] and the UML artifact of a meta resource class with its attributes as defined in Section III. The meta resource relates directly to FHIR resources by the participation relation, which models the collection of participating resources. Furthermore, the reference relation which models the *referenceStructure* property relates the meta resource to a ResourceTuple class, which combines two FHIR resources into a directed pair. For this use case, the *from* resource in the tuple references the *to* resource. A similar application is modeled for the composition relation, which models the *compositionStructure* property of a meta resource. Here the from resource contains the to resource of the tuple.

B. Domain Resources

The abstract resource DomainResource is the second facet we explore and is a foundation resource that provides common functionality to resources from which they were derived [11]. As an abstract resource, it is not directly instantiated, but because all of the standardized resources (except for Bundle, Parameter, and Binary) extend the DomainResource, its functionality is broadly available in the resource model. There is a property to a DomainResource that initially appears highly useful for including meta resources in the FHIR framework: the option to specify additional *contained* resources within a resource that closely relates to the grouping nature of meta resources. Since the *contained* property allows a designer to assemble other resources within a given resource (which then becomes a container), it provides functionality that fundamentally could express participating resources. Based on this, a possible approach would be to define a meta resource as a resource derived from DomainResource as illustrated in Figure 5. In the figure, the inheritance relation of DomainResource and Resource, the contained property, and a suggested extension relation of the MetaResource in the upper left would make the *contained* property available to all of the meta resources.

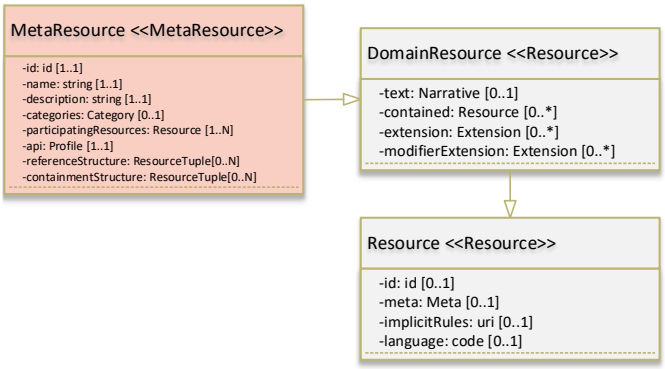


Figure 5. DomainResource Inheritance Structure.

However, closer inspection shows some significant limitations to this property and its application within the current version of the framework, limiting its value for incorporating meta resources into the standard. The resources referenced by the contained property are conceptually only existent in the scope of the container resource [11] and can neither be referred to without the container nor be passed around in their own transactions. This prohibits the idea of disassembling and partially using meta resources where they cannot be understood by plain FHIR systems and poses a problem to implement the reference and containment structures since they cannot reliably reference the contained elements without them having an own identity. Finally, contained resources may not contain further nested resources; therefore, a meta resource based on this property could not assemble arbitrary participating FHIR resources but would need to explicitly prevent them from nesting any data, which would harm the expressiveness of the model. As a result, the DomainResource functionality alone is not suitable to model meta resources within the FHIR standard.

C. FHIR Modules

The FHIR specification is organized in different functional areas to provide high-level guidance to implementers of the standard to make appropriate decisions on which parts of FHIR are required to model a particular healthcare process for solving a specific problem. These areas are reflected by *FHIR modules* [12] which is the third facet that we examine and aims to provide organizational and contextual information about the parts of the standard they cover. Towards this goal, modules are utilized to define: a *scope and index*, *use cases*, *security and privacy information*, and a *roadmap* for their content. The *scope and*

 ISBN: 978-1-61208-916-4

index provide developers with an introduction to the module, which consists of a textual description of the conceptual goals of the module and the envisioned functional range of the module. Furthermore, scope and index also summarize the technical content (i.e., the resources) of the module. The included *use cases* provide guidance and examples on using the contents of the module for implementing solutions. The *security and privacy information* component of the modules highlights areas of special caution and overall consideration for securing data expressed by the module. With FHIR still evolving as a standard, the roadmap provides information on the state of the evolution in maturity of the module's contents with respect to the overall standard.

As shown in Figure 6 [12], modules are classified into multiple levels, which offer a comprehensive and goal-focused view of FHIR information architecture. Levels 1 and 2 provide the technological foundation starting at low-level data types and exchange formats and extending to implementation basics and external specifications. Levels 3 and 4 contain parts of the standard used to model the healthcare domain's content and processes. Finally, Level 5 provides the means for reasoning over the information recorded and exchanged in the lower levels. Again, concepts from higher levels logically depend on the concepts introduced in the lower levels.

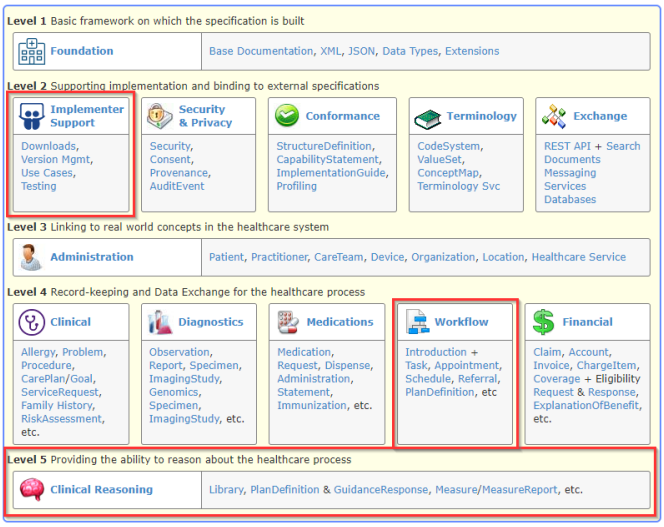


Figure 6. Organization of FHIR in Modules.

Note that modules are not necessarily mutually exclusive in their content. To illustrate this, consider that the PlanDefinition resource is used to define executable plans such as a chemotherapy regimen which means that it is both conceptually part of the workflow module in Level 4 as well as of the clinical reasoning module in Level 5. Similarly, the meta resources introduced by this proposal crosscut the module classification and can be primarily assigned to multiple levels specifically: the implementer support module in Level 2; the workflows module in Level 4; and, to some extent, to the clinical reasoning module in Level 5. Fundamentally, a meta resource (as with a design pattern) intends to simplify the development process of FHIR applications and eliminate repetitive decision processes by providing a specific framework for resource assembly. At the schema level, this is initially unrelated to the actual content of the resources and therefore fits into the implementer support module (at the technological foundation provided by Level 1 and Level 2). However, the composition of resources as part of the meta resource naturally allows for simplified and reusable modeling of generic workflows; therefore, actual instances of meta resource would be used to improve the workflow module in Level 4. Finally, the management of meta resources as artifacts that bundle lower-level resources, enrich them with additional knowledge about their relationships, and are provided as an advanced toolset in the standard would be classified into the clinical reasoning module in Level 5. Figure 7 illustrates the selected FHIR modules of the module architecture extended by meta resources (reduced to just the affected modules).

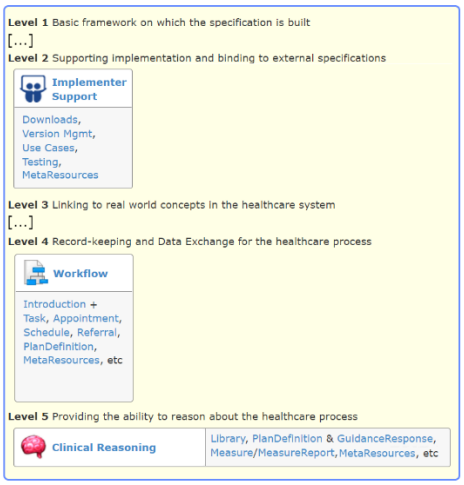


Figure 7. FHIR Modules extended by Meta Resources.

D. *Information Architecture*

Meta resources can be introduced into the FHIR standard as part of the information architecture, the fourth facet we explore, depicted as the composition framework in Figure 8 [9]. The resource contextualization Layer 6 of the architecture is designated to adapt FHIR resources to a specific environment. In contrast to the other layers, the information architecture serves a special purpose, as it does not contain resources but represents concepts that extend, constrain, add additional attributes, or provide meta information to the resources that are assigned to Layer 1 (Foundation Resources) through Layer 5 (Specialized Resources). Figure 8 illustrates a suggested extension of Layer 6 (hosting FHIR profiles and graphs) by the meta resource concept.

Profiles are used to tailor the FHIR specification to a specific environment. They achieve this by providing rules that allow a designer to restrict or extend the specification. In terms of resource attributes, profiles offer the ability to either disallow the use of specific attributes or add additional attributes to resources. An example use-case for this functionality is to facilitate the enforcement of adequate terminologies for a given domain. Profiles also allow the extension/restriction of the API that a FHIR server supports, therefore enabling an implementer to add custom services to the rest interface. In summary, profiles allow the modification of the data model of the specification and extend the supported means of communication, which is fundamental for supporting our meta resource extension.

 ISBN: 978-1-61208-916-4

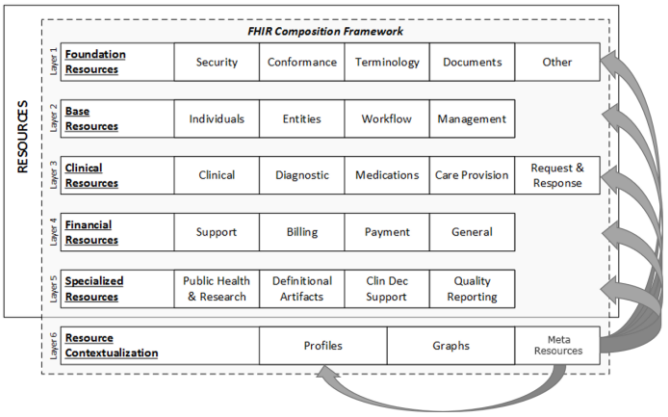


Figure 8. Information Architecture in FHIR.

For example, consider the mapping of the components of a meta resource serving for a medication reconciliation workflow to the relevant framework layers as displayed in Figure 9, where the box in the lower-left labeled MedRec Meta Resource indicates the system borders of the meta resources. The implementation of such a medication reconciliation meta resource (see Section III.A again) requires at least six FHIR resources: Medication, MedicationStatement, and DetectedIssue resources (*Clinical* resources in Layer 3) for modeling the medication information; the Patient and Practitioner resources (*Individuals* resources in Layer 2) for representing the involved individuals; and the Endpoint resource (*Entities* resources in Layer 2) to identify and address participating HER systems. Finally, the functionality of a specific MedRec FHIR profile is required to describe the APIs functionality for interacting with the meta resource (Layer 6).

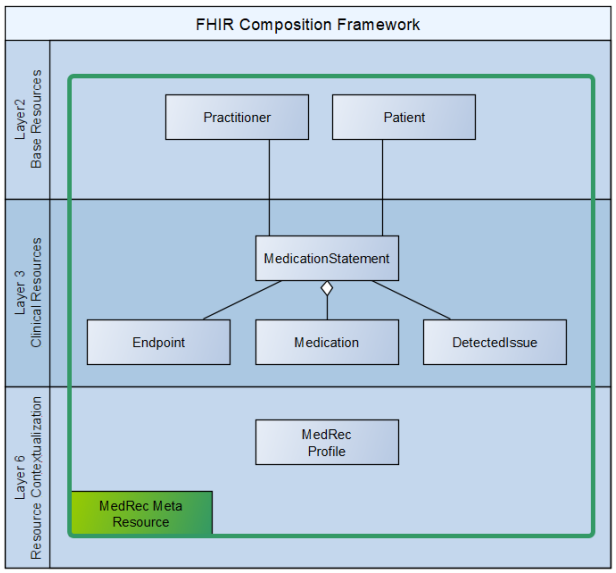


Figure 9. MedRec Meta Resource with Layer Associations.

E. Composition of Library and ActivityDefiniton Resources

In addition to Bundle resources, the FHIR standard offers further resources that can define collections of other resources, such as the Library resource and the ActivityDefinition resource, the fifth facet we explore, that can be combined to model an advanced level of resource composition. The FHIR Library resources are general purpose containers for expressing and sharing clinical knowledge assets independently of a particular patient [13]. The content of a Library resource (i.e., the *knowledge asset*) may consist of arbitrary non-FHIR data (e.g., the XML schema for an information model) but can equally consist of nested FHIR resources. Furthermore, Library resources also allow the definition of *parameter* properties for enforcing prerequisites to using the library's functionality. In this function, Library resources can provide some degree of composition of FHIR resources as required for the proposed meta resource.

The FHIR ActivityDefinition resource describes an activity in a sharable and potentially machine-consumable form [14], which can then be utilized to define parts of a workflow, to describe a protocol, or to create a catalog of activities. This functionality is consistent with our objective of realizing meta resource as a reusable, shareable, design pattern-like component. An ActivityDefinition is intended to be interpreted by humans but can also be filled with enough structured data to instantiate resources that reflect the modeled information. The combined use of Library resources and ActivityDefinition resources has the potential for utilization to represent a significant portion of the meta resource concept. The structure of this approach is illustrated in Figure 10, which shows the realization of meta resources together with a supporting ParticipantTuple structure (both extensions are at the bottom of the figure) through the extension of a Library resource that references an Activity resource definition (Library and ActivityDefinition are reduced in the diagram to the attributes relevant for the discussion).

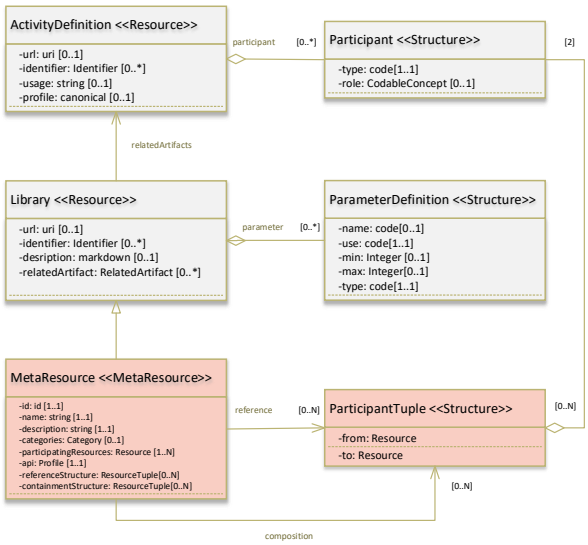


Figure 10. Composition of Library and ActivityDefinition.

The *name* and *description* elements of the meta resource can be realized by the respective *description* and *usage* attributes in Library and ActivityDefinition resources. The participating resources are defined by the parameter relation of the Library to a set of ParameterDefinition structures. Among a series of other types, ParameterDefinitions can be of type Resource, and therefore can encode the knowledge about required resource types for a meta resource. The Library resource then references

an ActivityDefinition, which in turn groups the instances of the participating resources as Participant structures and provides the necessary FHIR profile for interacting with the resource. As a result, the meta resource extension is only left with defining the *referenceStructure* and the *compositionStructure*, which can be achieved by introducing a directed ParticipantTuple relation between Participant instances (in the same fashion as suggested for ResourceTuples in Section IV.A)

V. CONCLUSION

This paper has presented the extension of the FHIR standard with the concept of a meta resource for design pattern-like capabilities and provided an examination of the alternative ways that meta resources can be positioned within five different facets of the FHIR standard. To present our work, we reviewed the background on medication reconciliation concepts and FHIR that we use for examples in Sections III & IV. We briefly reviewed the formal model for meta resources and the algorithmic translation of meta resources to FHIR bundles in Section III. In Section IV, we examined five facets of FHIR as alternative approaches for integrating meta resources into the FHIR standard. Specifically, we: explored meta resources as an extension to the FHIR UML model; assessed the relationship of the abstract DomainResource to meta resources; examined FHIR modules and the organizational and contextual information they could provide for meta resource integration; included meta resources into the FHIR information architecture; and considered meta resources as a composition of Library resources and ActivityDefiniton resources. Overall, we believe that meta resources offer an important design pattern-like capability to expand the FHIR standard and promote a much higher level of abstraction.

Ongoing work focuses on several different areas. To fully determine the benefits from using the meta resource concept, we intend to work with the team of the medication reconciliation application to assist in reformulating their usage of FHIR resources into a MedicationReconciliation meta resource as given in Section III. Another area is to fully formalize the appropriate notation so that the meta resource can fully fit into the defined FHIR standard by using only predefined FHIR conventions. Finally, we will explore generation options for FHIR facets introduced in Section IV in addition to the discussed bundle generation.

REFERENCES

[1] HL7 International. *FHIR Overview*. [Online]. Available from: https://www.hl7.org/fhir/overview.html 2021.09.08

[2] Office of the National Coordinator for Health Information Technology. *ONC's Cures Act Final Rule*. [Online]. Available from https://www.healthit.gov/curesrule/ 2021.09.08

[3] T. B. Ziminski, S. A. Demurjian, E. Sanzi, and T. Agresta, "Toward Integrating Healthcare Data and Systems: A Study of Architectural Alternatives," in Maximizing Healthcare Delivery and Management through Technology Integration, Iyamu, T., Tatnall, D., Eds. IGI Global, pp. 270-304, 2015

[4] T. B. Ziminski, S. A. Demurjian, E. Sanzi, M. Baihan, and T. Agresta, "Chapter 14: An Architectural Solution for Health Information Exchange," in Virtual and Mobile Healthcare: Breakthroughs in Research and Practice, IGI Global, pp. 283-327, 2020.

[5] T. B. Ziminski, S. A. Demurjian, and T. Agresta, "Extending the Fast Healthcare Interoperability Resources (FHIR) with Meta Resources," Proceedings of the 16th International Conference on Software Technologies, ICSOFT 2021, July 2021, in press.

[6] E. Gamma, *Design patterns: elements of reusable object-oriented software*, Addison-Wesley Professional, 1995.

[7] P. Kuchana, *Software architecture design patterns in Java*, CRC Press, 2004.

[8] J. A. Zachman. *Enterprise Architecture Defined: Architecture Abstractions*. [Online]. Available from https://www.zachman.com/resources/zblog/item/enterprise-architecture-defined-architecture-abstractions 2021.09.08

[9] HL7 International. *FHIR Architecture*. [Online]. Available from: https://www.hl7.org/fhir/overview-arch.html 2021.09.08

[10] HL7 International. *FHIR Specification: Base Resource Definitions*. [Online]. Available from: http://hl7.org/fhir/resource.html 2021.09.08

[11] HL7 International. *FHIR Specification: DomainResource Resource*. [Online]. Available from: http://hl7.org/fhir/domainresource.html 2021.09.08

[12] HL7 International. *FHIR Modules*. [Online]. Available from: https://www.hl7.org/FHIR/modules.html 2021.09.08

[13] HL7 International. *FHIR Specification: Library Resource*. [Online]. Available from: http://hl7.org/fhir/library.html 2021.09.08

[14] HL7 International. *FHIR Specification:ActivityDefinition Resource*. [Online]. Available from: http://hl7.org/fhir/activitydefinition.html 2021.09.08

[15] T. Agresta et al., "A Mobile Health Application for Medication Reconciliation Using RxNorm and FHIR". Proc. of the Fifth Intl. Conf. on Informatics and Assistive Technologies for Health-Care, Medical Support and Wellbeing (HEALTHINFO 2020), pp. 1-10, Oct 2020.

[16] CT State Legislature. *Connecticut SB00217 | 2018 | General Assembly, LegiScan*. [Online] Available from: https://legiscan.com/CT/bill/SB00217/2018 2021.09.08

[17] OHS Connecticut. *Work Group – Final Report*, pp 25-76 [Online]. Available from: https://portal.ct.gov/-/media/OHS/Health-IT-Advisory-Council/MRP/MRP_WG_FINAL_Recommendations.pdf 2021.09.08

[18] HL7 International. *FHIR Architecture*. [Online]. Available from: https://www.hl7.org/fhir/resourcelist.html 2021.09.08

[19] HL7 International. *FHIR Architecture*. [Online]. Available from: https://www.hl7.org/fhir/bundle.html 2021.09.08

 ISBN: 978-1-61208-916-4

Challenges on Real-World Skin Lesion Classification: Comparing Fine-tuning Strategies for Domain Adaptation using Deep Learning

Tudor Nedelcu, André Carreiro, Francisco Veiga and Maria João M. Vasconcelos
Fraunhofer Portugal AICOS
Porto, Portugal
email: maria.vasconcelos@fraunhofer.pt

***Abstract*—Skin lesion diagnosis is a challenging task even for experimented dermatologists. By using a computer-assisted diagnostic tool, misdiagnosed skin lesions are likely to decrease. Deep neural networks have emerged in recent years due to the increased computational power and their generalization capacity for new data. The major drawback of training a network is that it requires large amounts of data, often difficult to obtain. In this work, we introduce a real-world dataset of single lesions cases composed by clinical images, particularly challenging due to image variations (scale, size, point of view, acquisition device) and data imbalance. To tackle these challenges, we propose a domain adaptation approach by pre-training on large, general, datasets, such as ImageNet and fine-tuning on public datasets of clinical dermatological images. This approach is also compared with one where the target dataset is enriched with samples of public datasets. The overall performance obtained for this real-world dataset was not ideal, with F1-scores below 45%. However, interesting conclusions could be drawn on how fine-tuning generally yields better aggregated results (marginal increase of F1-score), although some specific categories benefit from increased training samples in a merged dataset. These results pave the way for new strategies towards the real-world application of skin lesion classification models, moving forward from more controlled settings, where results are typically impressive, however not yet translatable into practice.**

***Index Terms*—Skin Lesion Classification; Clinical Images; Transfer Learning; Deep Learning**

I. Introduction

There has been a growing interest in Telemedicine and other Information and communications technology (ICT) solutions to improve efficiency [1] and ease the burden on health services, but a significant potential still lies unexplored. Major advances in automatic risk assessment of skin lesions through computer-processed imaging have been recently reported, but most of this work has been conducted solemnly at an academic level and mainly focused on specific parts of the problem. There is a shortage of systems that convert the differently acquired knowledge into reliable decision support tools. Creating an integrated tool with effective practical utility is, thus, critical.

The last decades have seen great improvements concerning computer vision applications for clinical decision support, especially when Machine Learning, and more recently, Deep Learning, came into the picture. In fact, Deep Learning architectures have taken several computer vision tasks to new heights, from which Convolutional Neural Networks (CNN) stand out [2]. The main reason for the popularity of these networks, compared to traditional methods, is that they automatically learn features from images of a specific domain, without any explicit feature engineering. Another reason for their success is the possibility to transfer knowledge acquired for a specific task (resulting in a pre-trained model), to model a different task [2]. Training a CNN [2] from scratch, where the model weights are randomly initialized, requires large amounts of images and repetitive adjustments to the network and its parameters to avoid overfitting the training data [3]. For skin lesion classification, large datasets are a difficult requirement to meet, as the number of images in publicly available datasets is small and expert annotations are expensive [4].

Transfer learning is an alternative to training from scratch that allows to initialize the weights of the layers of a network that we intend to train in a new domain, using weights from a similar network previously trained on data from a different domain. A common practice is to replace the final output layers, where the model decision actually takes place, and freeze all other layers. In this scenario, only the new layers are trained with data from the new domain, keeping the weights of the lower layers fixed (frozen) [5]. This technique is also known as fine-tuning, since only the top layers are trained. In general, the first layers in a CNN learn more generic features, while the last layers learn more specific ones. If both the previously learned and the new domains are similar, fine-tuning the top few layers can be enough. However, if the target domain is considerably different from the source domain (learned by the pre-trained model), we may need to fine-tune the lower layers as well.

In this work, we introduce a new real-world, challenging, dermatological dataset and propose different approaches for training a skin lesion classification model using Deep Neural Networks (DNN) [2]. Firstly, different network architectures are evaluated. Then, different fine-tuning strategies are compared with a baseline where the target dataset is enriched with samples from public datasets containing the same categories. Results are presented and discussed, concerning the impact of the training strategies in classifying individual skin lesion classes, ending with major conclusions and drawing future research lines.

 ISBN: 978-1-61208-916-4

This work integrates a larger project, DermAI, that aims to improve the existing Teledermatology processes between Primary Care Units and Dermatology Services in the Portuguese National Health Service (NHS) for skin lesions referral. Through the usage of Artificial Intelligence and Computer Vision, we envision two major goals: a) to support doctors in Primary Care Units through the development of a mobile app that fosters image acquisition standardization [6] and b) to assist dermatologists in the referral process to book specialist consultations in the Hospital through the adequate prioritization of cases. Improving dermatology consultations' prioritization is particularly relevant in the Portuguese scenario, due to the lack of specialists in the NHS and long waiting lists for this type of consultation. In this research we focus on the second goal of the project towards cases prioritization, firstly on skin lesion classification.

This paper is structured as follows: Section 1 presents the motivation and objectives of this work; Section 2 summarizes background and related work found on the literature; Section 3 provides an overview of the methodology including datasets description, the network architectures studied as well as details on network training; in Section 4, the results and discussion are presented; and finally the conclusions and future work are drawn in Section 5.

II. Background and Related Work

In the last years, several approaches have been studied for using transfer learning in clinical applications. An important aspect of the works found in the literature on skin lesion classification is that the authors use well-established CNN architectures that have achieved excellent performance in large publicly available datasets, such as the ImageNet dataset [7], which consists of natural images of 1000 different categories. Fine-tuning on these pre-trained networks has shown outstanding results in new domains, even with smaller datasets [8] [9], improving both the performance and training times. Lopez et al. [10] proposed a method for skin lesion classification based on dermoscopic images, using a VGG16 network [11] trained for a binary malignant vs. benign classification task. The first four convolutional layers were the result of pre-training the network on ImageNet, whereas the remaining layers were fully trained with new images from the ISIC 2016 dataset [12]. This fine-tuning achieved a sensitivity of 78.66% and precision of 79.74%, which were significantly higher than the top evaluation results for the ISIC 2016 challenge (sensitivity of 50.70% and precision of 63.70%).

Gutman et al. [13] investigated the differences between training a model from scratch, compared with transfer learning and fine-tuning with application to dermatology domain, using EDRA dataset [14]. The model selected was VGG-M [15] with Support Vector Machine (SVM) as a classifier. For transfer learning and fine-tuning, the models were first trained on the Kaggle Retinopathy dataset [16] consisting of retinal images, ImageNet, or both (initially on ImageNet followed by former). The results showed that fine-tuning achieved better results than relying on frozen feature extraction. The models fine-tuned on Retinopathy, or both datasets, led to worse results than when fine-tuning with just ImageNet.

Kawahara et al. [17] trained a linear classifier on features extracted from an AlexNet network [18] pre-trained on ImageNet and fine-tuned on macroscopic images from the Dermofit dataset [19], which classified 10 different skin lesions with high accuracy.

Kawara et al. [20] further used two Inception-V3 [21] pre-trained networks on ImageNet (one for clinical images and one for dermoscopic images), for classification of diagnosis and skin lesion attributes prediction on the EDRA dataset. This work was extended by Nedelcu et al. [22], which pre-trained the networks on the ISIC2019 [23]–[25] dataset for dermoscopic images and Dermofit for clinical images, improving the classification performance.

Mahbod et al. [26] presented an ensemble technique using CNNs for skin lesion classification. The proposed method explores several CNN architectures (AlexNet, VGG16, ResNet-18, and ResNet-101 [27]) pre-trained on ImageNet and fine-tuned with dermoscopic images of skin lesions from the ISIC 2016 [12] and ISIC 2017 [23] datasets. Deep features are extracted from different layers from the different models that are then used to train an SVM. In addition, each pre-trained model is fine-tuned several times with different configurations, boosting the performance of a single architecture and the final results.

In summary, transfer learning has been extensively shown to improve results in different clinical domains, including skin lesion classification. However, this is typically achieved by pre-training in a large, general dataset such as ImageNet, and then fine-tuned on the target dataset, often comprised of dermoscopic images, or, less frequently, macroscopic images but very well standardized in regards to image quality and acquisition conditions. Our proposal to compare different training strategies in a very challenging real-world macroscopic dataset tries to overcome these limitations, including an experiment based on sequential fine-tuning resorting to other datasets, as was done by Gutman et al. [13]. Unfortunately, the results were worsened by using an intermediate dataset, presumably because it was drawn from an entirely different domain (retinal images).

III. Methodology

The main goal of this work is to perform skin lesion classification in a new private dataset, DermAI, consisting of macroscopic and anatomical images from single skin lesions, as shall be explained in Section III-A. Notwithstanding, training a successful classification model from scratch on this dataset is very challenging, especially considering the high number of classes, with a relatively low amount of data for each one (see Table I).

Thus, we explore the potential of fine-tuning on public datasets of related skin lesions, EDRA and Dermofit, further described in Section III-A, and assess how different combinations of these datasets with different characteristics influence the final performance.

 ISBN: 978-1-61208-916-4

As an alternative to transfer learning, we studied the impact of merging the available datasets in the classification performance. Figure 1 illustrates this process, noting that, because the classes considered in each dataset are different, we match DermAI categories with samples in the public datasets, discarding the unmatched categories from EDRA and Dermofit.

Figure 2 illustrates the different fine-tuning strategies under study. The baseline corresponds to training the model for classifying the 13 different skin lesions in DermAI dataset with weights trained from scratch. For the fine-tuning strategies, we fine-tune the DermAI dataset by pre-training on: just ImageNet (a); ImageNet, followed by EDRA (b); and finally, ImageNet, followed by EDRA and then by Dermofit (c). The next subsections describe the datasets used, the network architecture and parameters used for training and validation.

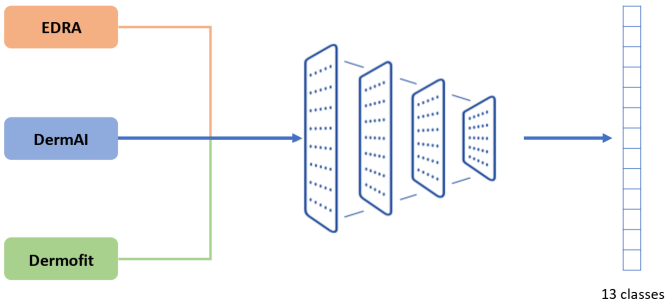


Figure 1. Training strategy based on a merged dataset. The goal is to classify the skin lesions available in DermAI. Thus, there is a previous mapping of the classes from the EDRA and Dermofit datasets, where unmatched class examples are dropped.

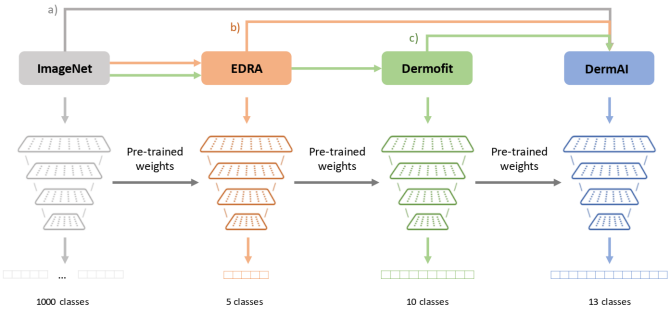


Figure 2. Training strategy based on transfer learning. The baseline considers only the training of DermAI from scratch, and then fine-tuning on DermAI is done sequentially and iteratively by pre-training on a) just ImageNet; b) ImageNet followed by EDRA; c) ImageNet followed by Edra and Dermofit.

A. Datasets

1) DermAI: With a larger goal of building a prioritization model for dermatological referrals, the authors had access to retrospective data from the Portuguese National Health System related to the referral requests from Local Health Care Units for the first Dermatology Hospital consultation. The cases correspond to requests that occurred between the implementation of this referral procedure for dermatological requests in 2013 to the end of February 2020, before the onset of the COVID-19 pandemic in Portugal. Since this data is retrospective it is not possible to publicly release it, due to confidentiality questions and impossibility to get consent from the patients.

After analyzing the available data, and together with a group of dermatologists, it was decided to proceed with a subset of 3430 cases corresponding to single lesions from 13 distinct differential diagnoses. The average age (and standard deviation) of the individuals correspond to 55.75 ± 22.21, and regarding the sex, there are 1422 Male cases, and 2042 Female instances. The distribution of cases in relation to the differential diagnosis provided by dermatologists is presented in Table I. Although teledermatology guidelines [28] recommend the acquisition of macroscopic images, in practice this is not always the case. Each case in the DermAI dataset has an associated clinical image: close-up/macroscopic, or anatomical. The dataset contains 3134 macroscopic images and 296 anatomical ones, as described in Table I. The type of images was selected by examining each type of image by the authors. In particular cases such as in lesions present in hands, arms, feet, or faces, it is difficult to differentiate between the macroscopic and anatomic labels since they look similar (close enough to evaluate the lesion, but wide enough to distinguish the anatomical site). Thus, it was decided to merge these modalities to train a DNN model. By merging the datasets, the number of samples is increased, which is beneficial for a smaller dataset such as this one.

Furthermore, as the authors had no access to biopsy results, in order to have increased confidence in the data, the authors have previously asked a set of dermatologists to review and validate a considerable subset of the test set.

2) EDRA: This is a public dataset [14] [20] that contains both clinical and dermoscopic images, as well as patient metadata. Clinical images are less standardized when compared with dermoscopic ones, meaning they are taken at various fields of view, and can also contain image artefacts such as rulers or other markers. Patient metadata includes other types of information, such as patient gender and lesion location. The dataset contains a total of 1101 cases distributed into 5 different categories: Seborrheic Keratosis, Miscellaneous (dermatofibroma, lentigo, melanosis, vascular lesion, miscellaneous), Nevus (blue, Clark, combined, congenital, dermal, recurrent, reed), Basal Cell Carcinoma and Melanoma. These classifications were assigned by a dermatologist, and the case distribution can be observed in Table II, where, for the scope of this work, only clinical images were considered. Moreover, some examples are presented in Figure 3.

3) Dermofit: Another public dataset is the Dermofit digital image database [19], which consists of 1300 high-quality color skin lesion images taken with standard cameras. The lesions belong to 10 different categories with 819 benign and 481 carcinogenic images, annotated in individual diagnostic classes (Table II). This dataset contains only close-up/macroscopic images and is the most standardized compared with the previous datasets, as it can be observed in Figure 3.

 ISBN: 978-1-61208-916-4

TABLE I
DERMAI DIFFERENTIAL DIAGNOSIS DATASET DISTRIBUTION.

Class	Differential diagnosis	Mac.	Anat.	Total
1 SebKer	Seborrheic Keratosis	1125	61	1186
2 ActKer	Actinic Keratosis	442	77	519
3 Nev	Nevus, Non-neoplastic	561	57	618
4 MolCont	Molluscum Contagiosum	50	21	71
5 Haem	Haemangioma	66	4	70
6 UncNeop	Neoplasm Unc. Behavior	233	13	246
7 Drmfib	Dermatofibroma	135	6	141
8 SLent	Solar Lentigo	45	3	48
9 PenFib	Pendulum Fibroma	99	16	115
10 VWart	Viral Warts	167	25	192
11 OtMalNeop	Other Malignant Neoplasm	108	8	116
12 BCC	Basal Cell Carcinoma	53	3	56
13 MM	Malignant Melanoma	50	2	52
	Total	3134	296	3430

TABLE II
EDRA AND DERMOFIT DIAGNOSIS DATASET DISTRIBUTION.

EDRA

Diagnosis	Total
Seborrheic Keratosis	45
Miscellaneous	97
Nevus	575
Basal Cell Carcinoma	42
Melanoma	252
Total	1101

Dermofit

Diagnosis	Total
Seborrheic Keratosis	257
Actinic Keratosis	45
Melanocytic Nevus	331
Haemangioma	97
Pyogenic Granuloma	24
Dermatofibroma	65
Intraepithelial Carcinoma	78
Squamous Cell Carcinoma	88
Basal Cell Carcinoma	239
Malignant Melanoma	76
Total	1300

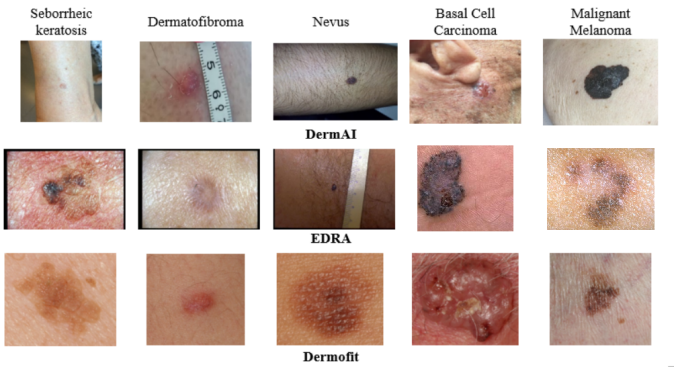


Figure 3. Illustrative examples of lesions from DermAI, EDRA and Dermofit datasets.

B. Merged Dataset

Given the unbalanced nature of the DermAI dataset, with some of the underrepresented classes being available in public datasets such as EDRA and Dermofit, we designed an experiment where such categories would be enriched by samples from those data sources. Examples of such classes, present in both EDRA and Dermofit, that can be joined into DermAI are Seborrheic Keratosis (SebKer), Nevus (Nev), Basal Cell Carcinoma (BCC), and Malignant Melanoma (MM). Additionally, from Dermofit we joined samples from Actinic Keratosis (ActKer) and Haemangioma (Haem). Furthermore, due to its low number of available samples, as well as its typical prioritization assessment in referrals to specialists, we considered examples from Pyogenic Granuloma into DermAI's class of Neoplasm of Uncertain Behavior (UncNeop), and images from Interepithelial Carcinoma and Squamous Cell Carcinoma into DermAI's category of Other Malignant Neoplasms (OtMalNeop). From the EDRA dataset, we extracted samples from the Miscellaneous class corresponding to Dermatofibroma (Drmfib) and Solar Lentigo (SLent) categories. The remaining classes in DermAI remained unchanged.

C. Network Architecture

Three different networks were studied, MobileNet-V2 [29], ResNet50 [30] and EfficientNet-B3 [31]. The MobileNet-V2 is one of the most adopted network for edge devices and is based on an inverted residual structure where the shortcut connections are between the thin bottleneck layers2 [29]. The ResNet50 is a residual network where the residual blocks make it easier to optimize, gaining accuracy from considerably increased depth [30]. The EfficientNet is a group of networks developed based on the network scaling (depths, width and resolution) [31]. An analysis performed on scaling ResNet and MobileNet networks has shown an increase on classification prediction on ImageNet. Although EfficientNet-B3 [31] was shown to surpass the other networks on ImageNet dataset, we also assess their performance on DermAI dataset for skin lesion classification.

On top of each architecture, a few layers are included for the final prediction. A fully connected (dense) layer is applied on top of the extracted feature map (EfficientNet-B3 $10 \times 10 \times 1536$, MobileNet-V2 $10 \times 10 \times 1280$, ResNet $10 \times 10 \times 2080$), generating a number of channels related to the number of classes to predict. For the DermAI dataset, the shape of this layer is $10 \times 10 \times 13$. For dimensionality reduction, the Global Average Pooling method is applied ($1 \times 1 \times 13$) since is known to reduce overfitting [32]. The final output is obtained by using the *softmax* activation function.

The input of the network consists of images of size 300×300. Since the images from the datasets have different shapes, we resize the images to the desired shape using the nearest neighbor method.

D. Network Training

The data is split into Train set and Test set with a ratio of 80:20 considering a stratified distribution of the classes.

The network is trained using the weights pre-trained on ImageNet (or from scratch with random initialization). The frozen block approach is adopted for better results [33]. Each block is trained for 3 epochs using a learning rate of 10^{-4} for the top layer and 10^{-5} for the rest of the blocks. Considering the EfficientNet-B3 architecture, 7 blocks are used for training (the classification block and other 6 modules). Adam is used as an optimizer, and the considered loss is the categorical cross-entropy.

A similar approach is followed for the MobileNet-V2 and ResNet50 networks (pre-trained on ImageNet), where

 ISBN: 978-1-61208-916-4

TABLE III
AVERAGE METRICS SCORE FOR DIFFERENT NETWORKS, AFTER PRE-TRAINING WITH IMAGENET (IN %).

Experiments	Number of Parameters	Average Accuracy	Weighted F1	Macro F1
MobileNet-V2	2.3M	14.43	15.91	9.59
ResNet50	23M	43.00	42.67	27.07
EfficientNet-B3	12M	42.71	44.04	28.65

the difference lies in the block mapping since the network architectures are different. For ResNet50, the classification block and 4 modules are considered, and for MobileNet-V2 the last 11 blocks. The blocks from MobileNet-V2 are grouped because of the residual connections as: block A - block 16; block B - block 13, 14, 15; block C - block 10, 11, 12; block D - block 6, 7, 8, 9.

To mitigate possible overfitting issues due to imbalanced data, we considered stratified batches, where the batch size was chosen to match the number of classes for each data set (5, 10, and 13 for EDRA, Dermofit, and DermAI, respectively). This results in oversampling of the classes with fewer examples. Additionally, we augmented the training data using simple techniques: rotation in the range of $[1, 30]$ degrees, horizontal flip, zooming in the range $[0, 0.2]$, width shift in the range $[0, 0.1]$, and brightness in the range $[0.2, 0.8]$.

IV. RESULTS AND DISCUSSION

The average metrics obtained for the three networks tested (with weights from ImageNet and fine-tuned on DermAI dataset) are presented in Table III. One can observe that MobileNet-V2 is performing poorly, with the network failing to provide acceptable performances. The results obtained using ResNet50 are similar to EfficientNet-B3, although more parameters are used. EfficientNet-B3 uses approximately 12 million parameters, whereas the ResNet50 uses almost double that number (23 million). Therefore, we chose to proceed with the experiments considering the EfficientNet-B3 network.

Table IV summarizes the overall aggregated results for the different considered approaches based on the chosen EfficientNet-B3 network architecture: training from scratch (o), using a merged version of the dataset (x), or different fine-tuning strategies (a-c).

The first experiment was to train the target dataset (DermAI) from scratch, to assess the importance of using pre-trained weights, even in general domains like ImageNet. The very poor results (Accuracy under 15% and F1-score macro under 3%) confirm this, and although we do not show the confusion matrix due to space constraints, it was observed a clear bias towards classifying most samples as Nevus (Nev) and Haemangioma (Haem), the second most and fourth least represented classes.

On the other hand, using models pre-trained on ImageNet drastically improved the results, even though these are still under what is expected from a clinical decision support tool in production. Comparing the averaged metrics, we can see that using a merged version of the training dataset (merging samples from EDRA and Dermofit into DermAI, where a match could be found between classes), returned a slightly lower accuracy and F1-score (weighted and macro) than when fine-tuning is used, be it just on DermAI, or sequentially with EDRA and Dermofit. Nonetheless, using a merged dataset returned some interesting results for particular categories, as can be seen in Table V. For instance, for the Haemangioma class, this approach was the one correctly classifying more samples, which may indicate that in some specific cases, there are discriminative features that can be more easily learned in the same learning process, although getting lost in the iterative process of sequential fine-tuning.

The results for fine-tuning strategies (a), (b), and (c) can be found on Tables VI to VIII. Through the analysis of the previous metrics and confusion matrices (where the predicted labels are on the abscissa and true labels are on the ordinate), comparing the different fine-tuning strategies, interesting findings can be highlighted for the individual categories of skin lesions.

TABLE IV
AVERAGE METRICS SCORE FOR DIFFERENT EXPERIMENTS (IN %).

Experiments	Aver. Acc.	Weight. F1	Macro F1
o) Training from scratch	14.28	5.52	2.53
x) Pre-train. ImageNet, merged Dataset	42.56	43.34	28.60
a) Pre-train. ImageNet	42.71	44.04	28.65
b) Pre-train. ImageNet and EDRA	43.73	44.17	30.09
c) Pre-train. ImageNet, EDRA, Dermofit	43.44	44.41	28.80

Regarding Seborrheic Keratosis (SebKer), which is the most represented class in the DermAI dataset, we report an F1-score of approximately 56% for pre-training on just ImageNet (a) and marginally higher for fine-tuning on EDRA (b) and additionally on Dermofit (c), with (b) and (c) revealing slightly lower sensitivity and higher precision. This class is well represented in the three considered datasets, and as expected, its classification was improved through more fine-tuning, although some misclassifications still happen, especially with the categories of Actinic Keratosis (ActKer), Nevus (Nev), and Neoplasm of Uncertain Behavior (UncNeop).

For the Actinic Keratosis (ActKer), also present in Dermofit but not in EDRA, the model returned F1-scores of approximately 52%, 56%, and 54% for strategies a), b) and c), respectively. This increase is due to the increase in sensitivity in strategies b) and c). This is especially interesting for b), given that this category is absent from the EDRA dataset, which might indicate that other categories from that dataset may present similar features which help to discriminate Actinic Keratosis. Typical misclassifications of this class include Seborrheic Keratosis (SebKer), Other Malignant Neoplasms (OtMalNeop), and Neoplasm of Uncertain Behavior (UncNeop).

Concerning the Nevus class (Nev), the second most represented category in DermAI and present in the 3 datasets (although EDRA comprises Melanocytic Nevus), the obtained F1-score results were very similar for all fine-tuning strategies

 ISBN: 978-1-61208-916-4

TABLE V
RESULTING METRICS WHEN PRE-TRAINING WITH IMAGENET AND FINE-TUNING ON MERGED DATASET, AND CORRESPONDING CONFUSION MATRIX.

Classes	Sens.	Prec.	F1
1 SebKer	53.59	66.84	59.48
2 ActKer	63.46	57.89	60.55
3 Nev	18.55	56.10	27.88
4 MolCont	42.86	27.27	33.33
5 Haem	21.43	21.43	21.43
6 UncNeop	18.37	11.11	13.85
7 Drmfib	53.57	31.25	39.47
8 SLent	10.00	8.33	9.09
9 PenFib	43.48	33.33	37.74
10 VWart	66.67	38.81	49.06
11 OtMalNeop	26.09	16.22	20.00
12 BCC	0.00	0.00	0.00
13 MM	0.00	0.00	0.00

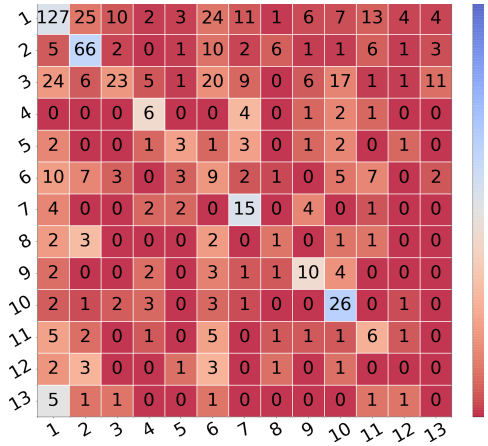
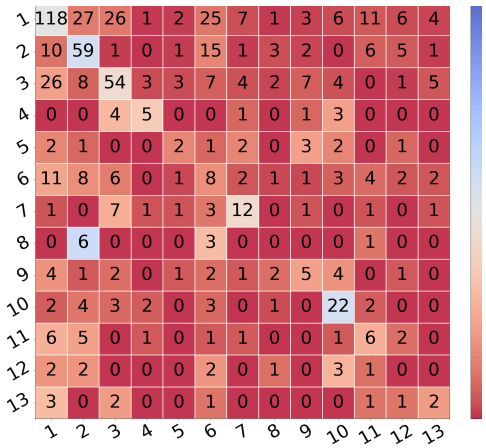


TABLE VI
RESULTING METRICS WHEN PRE-TRAINING WITH IMAGENET AND FINE-TUNING ON DERMAI, AND CORRESPONDING CONFUSION MATRIX.

Classes	Sens.	Prec.	F1
1 SebKer	49.79	63.78	55.92
2 ActKer	56.73	48.76	52.44
3 Nev	43.55	51.43	47.16
4 MolCont	35.71	38.46	37.04
5 Haem	14.29	18.18	16.00
6 UncNeop	16.33	11.27	13.33
7 Drmfib	42.86	38.71	40.68
8 SLent	0.00	0.00	0.00
9 PenFib	21.74	21.74	21.74
10 VWart	56.41	45.83	50.57
11 OtMalNeop	26.09	18.18	21.43
12 BCC	0.00	0.00	0.00
13 MM	20.00	13.33	16.00



(47%, 47%, and 48%, for a), b), c), respectively). However, looking at the sensitivity scores (44%, 46%, 48%), we can see that fine-tuning with additional dermatological datasets where examples of that class were present, considerably improved this important metric. Misclassifications for this category are biased towards Sebhorreic keratosis (SebKer), Neoplasm of Uncertain Behavior (UncNeop), and Dermatofibroma (Drmfib).

Analyzing the *Molluscum Contagiosum* class (absent from EDRA and Dermofit), we can observe F1-scores of approximately 37%, 43% and 38% for strategies a), b) and c). As for ActKer, fine-tuning with data with similar features, even though for different skin lesion categories, helps to better generalize when the number of samples is low, as is the case for this class.

A challenging category is Haemangioma (Haem), given the low amount of data. The F1-scores of 16%, 9% and 15% for strategies a), b), and c), does not support the advantage of using fine-tuning in this particular category, even though it is also present in Dermofit. Analyzing the erroneous classifications, these were more or less evenly distributed among different classes, such as Nevus, Pendulum Fibroma, Dermatofibroma, Sebhorreic Keratosis, and Viral Warts. To better understand these results, we analyzed these images more closely, concluding that, besides a low number of samples, they are very different amongst themselves (e.g., different body regions), making the model task more difficult.

Regarding the Dermatofibroma (Drmfib) category, which also exists in Dermofit and EDRA, the model returned F1-scores of 41%, 36%, and 33% for the three fine-tuning strategies (a), b), c), respectively). Although the differences seem considerable, given the small number of test cases, this translates to a difference of two correctly classified cases between a) and b), and one case between a) and c). Most misclassifications classify the samples as Nevus, and on a lesser extent with Neoplasm of uncertain behavior.

For Solar Lentigo (SLent), which has very few samples (under 50), and only exists in EDRA within the Miscellaneous class, returned poor results, as can be seen by the F1-scores of 0%, 17% and 0% for strategies a), b) and c), respectively. Again, as the difference might seem considerable at first, this corresponds to 0, 2, and 0 correctly classified samples in the test set. This category is expectedly difficult due to its low availability, with the model misclassifying these samples mostly with Actinic and Seborrheic Keratoses.

Concerning Pendulum Fibroma (PendFib), a class only present in DermAI, the results for F1-score for the fine-tuning strategies a), b), c) were 22%, 26%, and 24%, respectively. Even if only marginally, and given the low number of available images, pre-training seems to improve the results, especially

 ISBN: 978-1-61208-916-4

TABLE VII
RESULTING METRICS WHEN PRE-TRAINING WITH IMAGENET AND FINE-TUNING ON EDRA AND DERMAI, AND CORRESPONDING CONFUSION MATRIX.

Classes	Sens.	Prec.	F1
1 SebKer	47.68	69.75	56.64
2 ActKer	67.31	47.30	55.56
3 Nev	45.97	48.31	47.11
4 MolCont	42.86	42.86	42.86
5 Haem	7.14	11.11	8.70
6 UncNeop	4.08	5.13	4.55
7 Drmfib	50.00	28.00	35.90
8 SLent	20.00	15.38	17.39
9 PenFib	26.09	25.00	25.53
10 VWart	53.85	42.86	47.73
11 OtMalNeop	17.39	12.50	14.55
12 BCC	9.09	6.25	7.41
13 MM	30.00	25.00	27.27

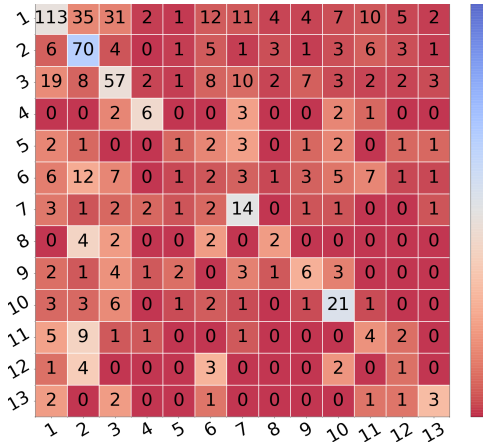
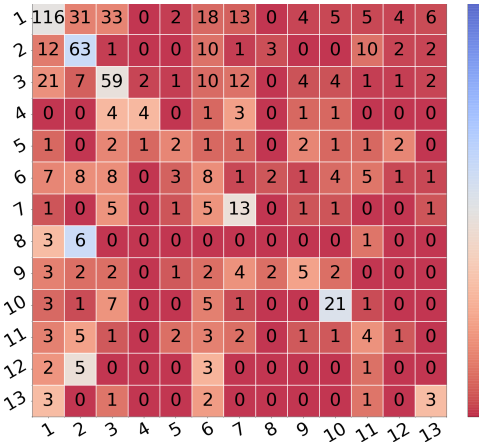


TABLE VIII
RESULTING METRICS WHEN PRE-TRAINING WITH IMAGENET AND FINE-TUNING ON EDRA, DERMOFIT, AND DERMAI, AS WELL AS THE CORRESPONDING CONFUSION MATRIX.

Classes	Sens.	Prec.	F1
1 SebKer	48.95	66.29	56.31
2 ActKer	60.58	49.22	54.31
3 Nev	47.58	47.97	47.77
4 MolCont	28.57	57.14	38.10
5 Haem	14.29	16.67	15.38
6 UncNeop	16.33	11.76	13.68
7 Drmfib	46.43	25.49	32.91
8 SLent	0.00	0.00	0.00
9 PenFib	21.74	26.32	23.81
10 VWart	53.85	52.50	53.16
11 OtMalNeop	17.39	13.33	15.09
12 BCC	0.00	0.00	0.00
13 MM	30.00	20.00	24.00



on sensitivity (22% in a) to 26% in b)). Misclassifications for this class do not reveal a clear trend, spanning several different categories.

The category of Viral Warts (VWart) is also present only in DermAI. The F1-score results of 51%, 48%, and 53% for strategies a), b), c), respectively reveal that this is one of the best-classified categories, and even though Dermofit does not present this category, pre-training with this dataset marginally improves the performance, especially regarding the precision (53%). Most misclassifications fall in the categories of Nevus, and Seborrheic and Actinic Keratoses, the most represented classes, which might support this observed bias, even though class-balancing methods were explored.

The category representing Neoplasm of Uncertain Behavior (UncNeop), as the name suggests, shows a higher variability within its samples. Thus, it is expected that this class is more challenging to classify, which is reflected in lower F1-scores of 13%, 5%, and 14% for fine-tuning strategies a), b), c), respectively. The drop for strategy b) might be a reflection of the under-representation of categories with similar features in the EDRA dataset. We recall that we have included the Pyogenic Granuloma cases into this category due to its low number of available samples and its clinical manifestations resulting in highly variable prioritization in case of referral, which is consistent with the UncNeop category. Since Dermofit also considers some examples of Pyogenic Granuloma, this might explain the marginal improvement of strategy c). Regarding misclassifications, these are mostly distributed among Seborrheic and Actinic Keratoses, Nevus, and Other Malignant Neoplasms, which is understandable as these are some of the classes with more diversity in their clinical manifestations.

Moving from the benign classes to the malignant ones, we start with Other Malignant Neoplasms (OtMalNeop), a broader category that includes the Intraepithelial Carcinoma and Squamous Cell Carcinoma, both present in the Dermofit dataset. Given the high case diversity, not supported by a sufficient number of data samples, lower results are somewhat expected, as observed in the F1-scores of approximately 21%, 15%, and 15% for the strategies a), b), and c), respectively. The drop in performance when pre-training with different datasets suggest that even though some of the learned features may be more general, this does not always imply better robustness for more naturally diverse categories, which can benefit from learning more freely on the available specific samples. Not surprisingly, both from a clinical perspective and due to their representativity in the dataset, most misclassifications fall in the Actinic and Sebhorreic Keratoses classes.

For Basal Cell Carcinoma (BCC), a class that is present in the three datasets, although with a low number of samples

 ISBN: 978-1-61208-916-4

in EDRA and DermAI, the model was only able to correctly classify one sample using fine-tuning strategy b) and none for the remaining strategies. BCC can have different biological and clinical manifestations [34], which by itself makes this class more complex to classify, especially with a lower sample availability. Nonetheless, the results motivated a deeper analysis of the corresponding images, and it was concluded that, for the DermAI dataset, the vast majority of examples in this class were anatomical images (mostly faces), whereas the images on EDRA and Dermofit were macroscopic images, centered on a single lesion. This mismatch between datasets, and even between partitions (train and test sets), together with the anticipated complexity, explains the poor performance for this category.

Finally, regarding the Malignant Melanoma (MM) class, it has examples in the three datasets, although more represented in EDRA. The classification performance resulted in F1-scores of 16%, 27%, and 24% for fine-tuning strategies a), b), and c), respectively. However, in absolute terms, given the low number of test samples (10), this corresponds to 2, 3, and 3 correctly classified instances. In terms of sensitivity, the values were 20%, 30%, and 30%, which hints at the positive impact of fine-tuning for this category where sensitivity is key. The low performance is mostly believed to be due low representation of this class, with only around 40 cases for training. As expected, also from a clinical point-of-view, misclassifications were biased towards Sebhorreic Keratosis and Nevus.

V. Conclusion and Future Work

The last decades have witnessed significant progress in what concerns computer-aided diagnosis for several domains, namely dermatology and skin lesion classification. However, most breakthroughs are limited to well-controlled environments, with data acquired in very specific conditions, putting almost all effort in model development and improvements thereof. Despite all the impressive results found in the literature, most systems rely on non-standard acquisition equipment, handled by professionals whose focus is far from ensuring data quality and standardization.

To address this gap in real-world applications, this paper proposes different strategies for training a real-world image dataset for skin lesion classification, comprised of retrospective data from the Portuguese National Health System - DermAI. It presents 13 different differential diagnostic categories and its images are very diverse concerning the acquisition settings, the field-of-view, and overall quality, making it a challenging dataset, especially when compared to publicly available ones, as EDRA and Dermofit.

Different methodologies for training a Deep Neural Network on this unbalanced dataset were studied. First, we evaluated different network architectures with available pre-trained weights on ImageNet to assess with which one to proceed for the following experiments: MobileNet-v2, Resnet50 and EfficientNet-B3. Considering the trade-off between trainable parameters (model complexity) and performance (based on F1-score), we chose to proceed with the EfficientNet-B3.

Following, a model was trained from scratch, using only DermAI image samples. Unsurprisingly, these results were very poor (Accuracy of 14% and macro F1-score of 3%), with the model overfitting for two of the classes: Nevus and Haemangioma. A different, straightforward approach, relied on merging samples from common classes in the public skin lesion datasets into DermAI. In opposition, we also propose a sequential fine-tuning pipeline where the target dataset is fine-tuned after pre-training the model iteratively with other datasets, from larger and more general sets (ImageNet) to smaller, similar domain, ones, like EDRA and Dermofit. One of the first conclusions is that pre-training, even with datasets as general as ImageNet, have a significant impact on the model performance (Accuracy over 40% and macro F1-scores over 28%), even though the results reflect the complexity of the DermAI dataset.

When comparing the use of a merged dataset (using common skin lesion classes) with the use of sequential fine-tuning, the conclusions are less clear. In general, sequential fine-tuning returns marginally higher aggregated results (a marginal increase of macro F1-score), although some specific categories benefit from increased training samples in the same learning step. This is especially evident for classes where the images show significant differences between the datasets (e.g., Haemangioma or Pendulum Fibroma), since the first pre-training might lead the learning process away from extracting features that are more discriminative for categories only represented in the target dataset. On the other hand, other categories may benefit from such an iterative learning process, where features learned in other datasets help to generalize examples in the target dataset (e.g., Nevus and Malignant Melanoma). Moreover, although the averaged metrics favor a pre-training on EDRA and not including Dermofit, for some categories the latter shows to be beneficial, such as for Haemangioma, Neoplasm of Uncertain Behavior, and Viral Warts.

Our results highlight the challenges of a real-world application of skin lesion classification models, highly dependent on the available data, especially concerning its amount, quality, and diversity. One major conclusion is that high-quality, low-cost, and portable acquisition systems assume paramount importance in building good training sets. Furthermore, as disease cases are hard to get, in order to explore existing data, data-centric techniques should be explored in future work to improve results. These may include simply other aspect-ratio preserving resizing methods, or more interestingly, lesion segmentation methods to help the model focus on the most critical regions of the image without removing all of the, also important, surrounding context. Additionally, the use of neural activation maps or other related explainable methods could help analyse specific errors that may inspire further image preprocessing steps. Concerning data merging versus sequential fine-tuning, this work opens the door for future research on exploring both approaches simultaneously. It is possible to merge common classes (especially if that category's image variance is high) and follow with a sequential fine-tuning process for missing or not-shared classes. Further

 ISBN: 978-1-61208-916-4

future work considers including meta-data in the classification models and investigating joint training of multiple tasks related to skin lesion classification since parameter-sharing could be a good alternative to sequential pre-training. Finally, hierarchical classification can also be explored in future experiments, where different layers are considered towards improving the final classification. For example, one can first classify a sample into Benign, Malignant, or Uncertain categories and then into the corresponding final differential diagnosis based on the first decision.

ACKNOWLEDGMENT

This work was done under the scope of project "DERM.AI: Usage of Artificial Intelligence to Power Teledermatological Screening", and supported by national funds through 'FCT—Foundation for Science and Technology, I.P.', with reference DSAIPA/AI/0031/2018.

REFERENCES

[1] D.-G. for Health and F. Safety, "Market study on telemedicine," tech. rep., European Commission, 2018.

[2] S. Srinivas, R. K. Sarvadevabhatla, K. R. Mopuri, N. Prabhu, S. S. Kruthiventi, and R. V. Babu, "A taxonomy of deep convolutional neural nets for computer vision," *Frontiers in Robotics and AI*, vol. 2, p. 36, 2016.

[3] E. Valle, M. Fornaciali, A. Menegola, J. Tavares, F. V. Bittencourt, L. T. Li, and S. Avila, "Data, depth, and design: Learning reliable models for skin lesion analysis," *Neurocomputing*, vol. 383, pp. 303–313, 2020.

[4] C. N. Vasconcelos and B. N. Vasconcelos, "Convolutional neural network committees for melanoma classification with classical and expert knowledge based image transforms data augmentation," *arXiv preprint arXiv:1702.07025*, 2017. [retrieved: June,2021].

[5] L. T. Thao and N. H. Quang, "Automatic skin lesion analysis towards melanoma detection," in *2017 21st Asia Pacific symposium on intelligent and evolutionary systems (IES)*, pp. 106–111, IEEE, 2017.

[6] D. Moreira, P. Alves, F. Veiga, L. Rosado, and M. J. M. Vasconcelos, "Automated mobile image acquisition of macroscopic dermatological lesions," in *HEALTHINF*, pp. 122–132, 2021.

[7] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*, pp. 248–255, Ieee, 2009.

[8] V. Pomponiu, H. Nejati, and N.-M. Cheung, "Deepmole: Deep neural networks for skin mole lesion classification," in *2016 IEEE International Conference on Image Processing (ICIP)*, pp. 2623–2627, IEEE, 2016.

[9] K. M. Hosny, M. A. Kassem, and M. M. Foaud, "Classification of skin lesions using transfer learning and augmentation with alex-net," *PloS one*, vol. 14, no. 5, p. e0217293, 2019.

[10] A. R. Lopez, X. Giro-i Nieto, J. Burdick, and O. Marques, "Skin lesion classification from dermoscopic images using deep learning techniques," in *2017 13th IASTED international conference on biomedical engineering (BioMed)*, pp. 49–54, IEEE, 2017.

[11] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014. [retrieved: June,2021].

[12] D. Gutman, N. C. Codella, E. Celebi, B. Helba, M. Marchetti, N. Mishra, and A. Halpern, "Skin lesion analysis toward melanoma detection: A challenge at the international symposium on biomedical imaging (isbi) 2016, hosted by the international skin imaging collaboration (isic)," *arXiv preprint arXiv:1605.01397*, 2016. [retrieved: June,2021].

[13] A. Menegola, M. Fornaciali, R. Pires, F. V. Bittencourt, S. Avila, and E. Valle, "Knowledge transfer for melanoma screening with deep learning," in *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, pp. 297–300, IEEE, 2017.

[14] G. Argenziano, H. Soyer, V. De Giorgi, D. Piccolo, P. Carli, M. Delfino, and Others, "Dermoscopy: a tutorial," *EDRA, Medical Publishing & New Media*, 2002.

[15] K. Chatfield, K. Simonyan, A. Vedaldi, and A. Zisserman, "Return of the devil in the details: Delving deep into convolutional nets," in *Proceedings of the British Machine Vision Conference*, BMVA Press, 2014.

[16] EyePACs, "Diabetic retinopathy detection kagle." https://www.kaggle.com/c/diabetic-retinopathy-detection/data. [retrieved: June,2021].

[17] J. Kawahara, A. BenTaieb, and G. Hamarneh, "Deep features to classify skin lesions," in *2016 IEEE 13th international symposium on biomedical imaging (ISBI)*, pp. 1397–1400, IEEE, 2016.

[18] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems*, vol. 2, pp. 1097—1105, 2012.

[19] L. Ballerini, R. B. Fisher, B. Aldridge, and J. Rees, "A color and texture based hierarchical k-nn approach to the classification of non-melanoma skin lesions," in *Color Medical Image Analysis*, pp. 63–86, Springer, 2013.

[20] J. Kawahara, S. Daneshvar, G. Argenziano, and G. Hamarneh, "Seven-point checklist and skin lesion classification using multitask multimodal neural nets," *IEEE journal of biomedical and health informatics*, vol. 23, no. 2, pp. 538–546, 2018.

[21] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2818–2826, 2016.

[22] T. Nedelcu, M. Vasconcelos, and A. Carreiro, "Multi-dataset training for skin lesion classification on multimodal and multitask deep learning," in *Proceedings of the 6th World Congress on Electrical Engineering and Computer Systems and Sciences (EECSS'20)*, pp. ICBES 120–1–8, 2020.

[23] N. C. Codella, D. Gutman, M. E. Celebi, B. Helba, M. A. Marchetti, S. W. Dusza, *et al.*, "Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (isbi), hosted by the international skin imaging collaboration (isic)," in *2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018)*, pp. 168–172, IEEE, 2018.

[24] P. Tschandl, C. Rosendahl, and H. Kittler, "The ham10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions," *Scientific data*, vol. 5, no. 1, pp. 1–9, 2018.

[25] M. Combalia, N. C. Codella, V. Rotemberg, B. Helba, V. Vilaplana, O. Reiter, *et al.*, "Bcn20000: Dermoscopic lesions in the wild," *arXiv preprint arXiv:1908.02288*, 2019. [retrieved: June,2021].

[26] A. Mahbod, G. Schaefer, I. Ellinger, R. Ecker, A. Pitiot, and C. Wang, "Fusing fine-tuned deep features for skin lesion classification," *Computerized Medical Imaging and Graphics*, vol. 71, pp. 19–29, 2019.

[27] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

[28] L. M. Abbott, R. Miller, M. Janda, H. Bennett, M. Taylor, C. Arnold, S. Shumack, H. P. Soyer, and L. J. Caffery, "Practice guidelines for teledermatology in australia," *Australasian Journal of Dermatology*, vol. 61, no. 3, pp. e293–e302, 2020.

[29] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4510–4520, 2018.

[30] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

[31] M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *International Conference on Machine Learning*, pp. 6105–6114, PMLR, 2019.

[32] M. Lin, Q. Chen, and S. Yan, "Network in network," *arXiv preprint arXiv:1312.4400*, 2013. [retrieved: June,2021].

[33] T. Shermin, S. W. Teng, M. Murshed, G. Lu, F. Sohel, and M. Paul, "Enhanced transfer learning with imagenet trained classification layer," in *Pacific-Rim Symposium on Image and Video Technology*, pp. 142–155, Springer, 2019.

[34] A. I. Rubin, E. H. Chen, and D. Ratner, "Basal-cell carcinoma," *New England Journal of Medicine*, vol. 353, no. 21, pp. 2262–2269, 2005.

 ISBN: 978-1-61208-916-4

A Blueprint Towards an Integrated Healthcare Information System Through Blockchain Technology

Ghassan Al-Sumaidaee*
ECE department
McGill University
Montreal, Canada
ghassan.al-sumaidaee@mail.mcgill.ca

Rami Alkhudary
LARGEPA
Université Paris II Panthéon-Assas
75231 Paris, France
rami.alkhudary@etudiants.u-paris2.fr

Zeljko Zilic
ECE department
McGill University
Montreal, Canada
zeljko.zilic@mcgill.ca

Pierre Féniès
LARGEPA, Université Paris II Panthéon-Assas
75231 Paris, France
pierre.fenies@u-paris2.fr

***Abstract*— Since the advent of the Internet, software development has continuously transformed the healthcare industry. Electronic medical or health records are prominent examples of healthcare data digitization. However, these records are vulnerable to cyber-attacks and data loss. Careful attention must be paid to data integrity, patient privacy, data management, and storage concerns. Blockchain technology is proposed in the literature to integrate healthcare information systems through a decentralized and unified network. Conceptual configurations of blockchain-based healthcare information systems are conquering the literature. Nevertheless, these studies lack empirical evidence. In response, we employ the design science research methodology and target the healthcare information system in Montreal, Canada. This paper presents an empirical research blueprint that proposes an integrated healthcare information system through blockchain technology.**

Keywords-blockchain; decentralized information system; healthcare supply chain; data management.

I. INTRODUCTION

Building an efficient and trusted healthcare information system is arduous, given the direct impact on patients and global health [1][2]. In addition, the very primitive tools in some medical supply chains continue to challenge healthcare data management. Counterfeit medicinal drugs are another consistent and growing challenge that impacts recipients and the costs associated. According to a joint study carried out by the European Union Intellectual Property Office (EUIPO) and the Organization for Economic Co-operation and Development (OECD), the total expenditure on trade in counterfeit medicines was €4.03 billion between 2014 and 2016 [3].

The academic literature suggests using blockchain in healthcare to enhance the security and privacy of centralized information systems. In addition to the financial industry, blockchain as a distributed ledger technology attracts academia and various industries. Blockchain is defined as a decentralized database or distributed ledger of digital records or financial transactions that are immutably registered according to a precise consensus mechanism [4]–[6]. Blockchain's ledger consists of concatenated blocks of information or digital assets using powerful and unbreakable cryptographic techniques to ensure the integrity and security of the registered data. The consensus mechanism for validating data in the blockchain ledger is a key element that distinguishes blockchain from other enterprise resource planning systems [7][8].

Digital records need to be validated to be registered in the blockchain ledger indefinitely. Otherwise, it is rejected. Small, medium, and large companies have been tested blockchain in the supply chain to ensure better financial, physical, and information flows [9]–[11]. The healthcare industry is no exception [12]–[14].

Blockchain applications in healthcare span multiple aspects and focus on data management, traceability, transparency, privacy, and security [15]. For example, tracking medications through blockchain within an integrated health information system would drastically reduce tampering attempts. In addition, blockchain could provide a reliable and secure environment for data exchange between healthcare actors and the government in real-time [16][17].

This research is motivated by the need to develop an integrated healthcare information system through blockchain. This is because the integration of healthcare stakeholders' information systems remains out of reach, raising many problems concerning patients' privacy and poor treatment of medical data. This research aims to build and test a blockchain-based platform that connects different healthcare stakeholders in Montreal, Canada.

First, our study explores the possibility of using the private blockchain model (i.e., Hyperledger Fabric) to integrate different healthcare information systems and facilitate their access to a unified platform. Second, we propose hosting the blockchain ledger either on the cloud (i.e., Amazon Web Services or Microsoft Azure Cloud) or servers owned by the healthcare network. Third, we build application programming interfaces and secure access to the platform.

This paper is organized as follows. Section II reviews the literature on blockchain applications in healthcare. Section III describes our methodology represented in the design science research approach. Section IV presents our preliminary findings. To finish with, Section V concludes with remarks.

II. LITERATURE REVIEW

Blockchain can be divided into three main models: public, private, and consortium [4][5][18][19]. In a public blockchain system, access is open to all users where registered information is transparent. In addition, users can participate in the validation process of the network's transactions. The

 ISBN: 978-1-61208-916-4

transactions in a private blockchain system may not be fully transparent. This is because one or more entities often control the network. Also, users who want to join the private blockchain network must get permission. A semi-decentralized blockchain, called a consortium blockchain, is managed by more than one entity, and all are involved in decision-making on the blockchain network [20]. It is worth noting that most studies in the literature do not recognize different blockchain models, nor do they consider the benefits associated with each model in practice.

McGhin et al. [21] highlight the need to secure electronic health and medical records by ensuring that only the authorized stakeholders have access to the records. Also, monitoring should be achieved to prevent the risks of possible tampering. The authors propose using blockchain to manage healthcare data better and mitigate the risk of data loss and cyber-attacks.

Benil and Jasper [22] propose a public auditing scheme for the healthcare industry called the Elliptical Curve Certificateless Aggregate Cryptography Signature-scheme (EC-ACS). The scheme suggests using blockchain to secure electronic healthcare records on the cloud.

In order to enhance patients' privacy and protect their medical records, Zhuang et al. [23] develop a private blockchain model that enables a patient-centric Health Information Exchange (HIE). The authors explain the strength of this technology in HIE and attempt to assess the feasibility of its adoption.

In a similar vein, Shi et al. [24] address the security and privacy policies of the Health Insurance Portability and Accountability Act (HIPAA). The authors show that blockchain can improve the healthcare supply chain's resilience and maintain its integrity.

III. METHODOLOGY

As previously mentioned, we explore blockchain in the Canadian healthcare system. We are in the process of building an integrated blockchain-based information system that facilitates and secures medical records and monitors COVID vaccination. To do so, we follow the design science research methodology [25][26]. Accordingly, we divide our study into five phases as shown in Figure *1*.

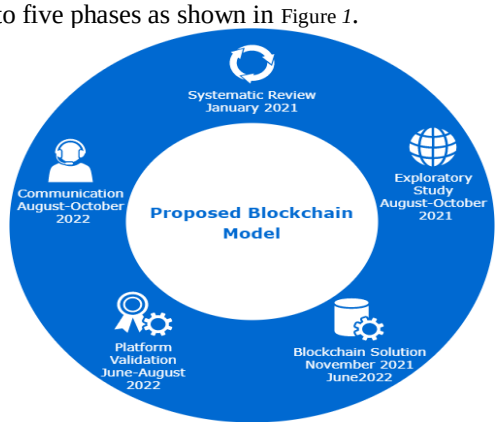


Figure 1. Progress of research phases.

In the first phase, we conduct a systematic review of the literature [27]–[29] to identify the problems that confront healthcare information systems. In the second phase, we conduct interviews with the McGill University Health Centre (MUHC) and its stakeholders to assess the literature claims. In the third phase, we build the technical solution: a blockchain-based healthcare information system to integrate different stakeholders' information systems and improve the integrity, privacy, and security of medical data. Other technologies (e.g., Internet of Things devices) will be considered to ensure data veracity. In the fourth phase, we test our system at the MUHC. In the fifth phase, we modify and approve the proposed system. Below is a brief explanation of the five phases.

A. *Phase 1: A systematic review of the literature (completed)*

In this phase, we review the literature on blockchain applications in the healthcare industry. We formulate the main propositions about the benefits that blockchain can bring to the healthcare industry and the challenges that it can resolve.

A set of inclusion and exclusion criteria was defined to collect and analyze the literature. First, we employed two academic search engines: Scopus and Business Source Ultimate. Many search strings were tested to assess the relevance and size of the literature. This procedure produced 277 academic papers. Second, we read each paper's title, abstract, keywords, and more text if needed to evaluate the eligibility of the collected studies. We only included papers that examined the use of blockchain in healthcare as a primary issue. This procedure minimized the number of articles to 142. Third, each article was coded to capture the blockchain's benefits and the challenges that it can resolve. Fourth, we carefully assessed the boundary conditions of each paper. Fifth, we highlighted the consensus algorithms proposed or studied in the literature for blockchain-based healthcare systems. Four main algorithms were identified: Proof-of-Work (PoW), Proof-of-Stake (PoS), Proof of Authority (PoA), and Practical Byzantine Fault Tolerance (PBFT).

B. *Phase 2: Exploratory study (occurring at the time of writing)*

After identifying blockchain benefits and the challenges that it can resolve in healthcare, we conduct several interviews with the MUHC to refine the literature's claims (questions are listed below). It is worth noting that we do not aim to identify the problems of the entire Canadian healthcare system. This is a daunting task that may take years. Our research is a step forward in this direction. More precisely, our unit of analysis is the MUHC and its stakeholders.

- What is the current design of the MUHC supply chain? Who are its main stakeholders?
- What type of communication systems are used by each stakeholder? What kind of data is exchanged? Where is the medical data hosted?
- Are there any supplementary technologies that need to be used with blockchain to ensure data veracity?

 ISBN: 978-1-61208-916-4

- What are the most apparent issues challenging the current communication systems?

In addition, we conduct a market study to evaluate the most recent blockchain solutions proposed in the industry for healthcare (occurring at the time of writing). This study aims to answer several questions (listed below) that will help us refine our blockchain system.

- What model of blockchain is used?
- What is the consensus mechanism to validate and register information on the blockchain ledger?
- Where is the medical data hosted?
- Who controls the blockchain network (if any)?
- What is the added value of the blockchain solution in healthcare? Is it measured or quantified?

C. *Phase 3: Building the blockchain solution (occurring at the time of writing)*

In this phase, we build (code) the blockchain database. As previously mentioned, we develop a private blockchain system on Hyperledger Fabric. Hyperledger is an umbrella project of open-source blockchains. It offers a high degree of flexibility and multiple ordering services, making it suitable for private sector collaborations. After building the technical solution, we ensure better control over the stakeholders' access.

Overall, the proposed system enables healthcare stakeholders to exchange medical data in real time. According to the network designed during the exploratory study, the information registered on the blockchain ledger would be transparent to all or some stakeholders. Figure 2 shows a sketch of the technical solution. The proposed system and its components will be refined after the completion of the exploratory study, which is still ongoing at the time of writing.

After considering this sector's high confidentiality and authenticity, the system's apparent goal is to integrate the MUHC's stakeholders' information systems with a unified and secure blockchain database.

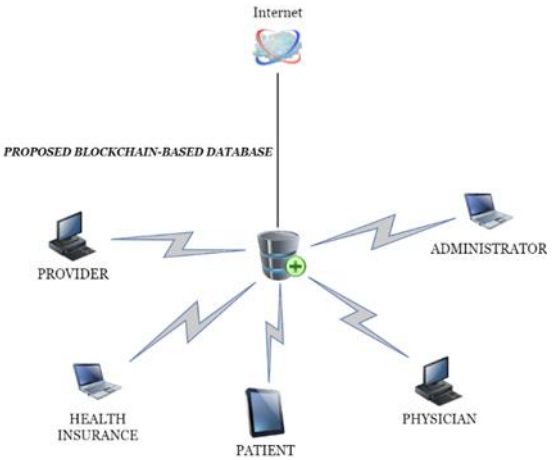


Figure 2. Blockchain unified database.

D. *Phase 4: Platform validation (to be completed during the first half of 2022)*

After building the system, another evaluation process will be conducted to assess and validate the proposed blockchain database. More precisely, the system will be subject to several test experiments at the MUHC. In addition, we interview experts, doctors, managers, information technology specialists working in the Canadian healthcare system.

E. *Phase 5: Communication and model evaluation (to be completed during the first half of 2022)*

In the fifth phase, if the results obtained are as expected, our system will be valid. Otherwise, we would take a step back to Phase 3 to make the necessary improvements. To proceed with this phase, we may need to conduct further interviews with the MUHC and its stakeholders to develop further or modify the system.

IV. PRELIMINARY FINDINGS

We have conducted a systematic review of the literature and analyzed the collected papers. Our exploratory study is occurring at the time of writing. Accordingly, we can report some of the preliminary findings.

Our review of blockchain applications in the healthcare industry revealed that many consensus algorithms are not suitable for healthcare. Many blockchain architectures and new supply chain configurations can emerge by changing the blockchain consensus mechanism.

Proof-of-Authority (PoA) and Practical Byzantine Fault Tolerance (PBFT) are examples of suitable mechanisms. PoW consumes an enormous amount of energy to validate and register information. Although the PoS is considered a solution for the energy problem associated with PoW, this algorithm does not allow healthcare stakeholders to register medical data flexibly. In addition, we categorized tens of blockchain benefits that blockchain can offer to healthcare, such as traceability, transparency, security, privacy, anonymity, interoperability, etcetera.

Our preliminary findings from the exploratory study highlight that current healthcare systems are highly fragmented. In other words, each stakeholder has its own information system. Accordingly, information is poorly communicated with the different stakeholders, which causes significant delays. In addition, medical data is not secure and vulnerable to data loss and cyber-attacks. These factors negatively impact data management and patients' satisfaction.

Although blockchain is proposed to solve these problems, this potential seems to be still conceptual. We found that many companies today (e.g., IBM) offer easy-to-use blockchain solutions for the healthcare industry. However, these solutions are standard and difficult to customize to fulfill specific needs.

 ISBN: 978-1-61208-916-4

V. CONCLUDING REMARKS

Research investigating the potential of blockchain in healthcare continues to grow, particularly concerning security, privacy, integrity, and interoperability. Most articles in the literature discuss the capacity of this technology to secure medical records. While this cannot be entirely dismissed, we believe that this has yet to be proven in practice.

Blockchain technology has proven successful and attracted the attention of several industries. However, there is still a considerable need to develop and test customized blockchain solutions in practice. In response, our proposed system is a step forward toward building a decentralized healthcare supply chain. Our system solves the integration problem highlighted in the healthcare industry through a design science research approach.

REFERENCES

[1] K. A. Clauson, E. A. Breeden, C. Davidson, and T. K. Mackey, "Leveraging Blockchain Technology to Enhance Supply Chain Management in Healthcare:: An exploration of challenges and opportunities in the health supply chain," *BHTY*, Mar. 2018, doi: 10.30953/bhty.v1.20.

[2] A. A. Abdellatif, A. Z. Al-Marridi, A. Mohamed, A. Erbad, C. F. Chiasserini, and A. Refaey, "ssHealth: Toward Secure, Blockchain-Enabled Healthcare Systems," *IEEE Network*, vol. 34, no. 4, pp. 312–319, Jul. 2020, doi: 10.1109/MNET.011.1900553.

[3] C. Boudot, "COUNTERFEIT MEDICINES: A GLOBAL THREAT HEIGHTENED BY COVID-19," Jun. 2020. [Online]. Available: https://servier.com/en/news/counterfeit-medicines-global-threat-covid19/

[4] R. Alkhudary, X. Brusset, and P. Fenies, "Blockchain in general management and economics: a systematic literature review," *EBR*, vol. 32, no. 4, pp. 765–783, Jul. 2020, doi: 10.1108/EBR-11-2019-0297.

[5] D. E. O'Leary, "Configuring blockchain architectures for transaction information in blockchain consortiums: The case of accounting and supply chain systems," *Intell Sys Acc Fin Mgmt*, vol. 24, no. 4, pp. 138–147, Oct. 2017, doi: 10.1002/isaf.1417.

[6] M. M. Queiroz, R. Telles, and S. H. Bonilla, "Blockchain and supply chain management integration: a systematic review of the literature," *SCM*, vol. 25, no. 2, pp. 241–254, Aug. 2019, doi: 10.1108/SCM-03-2018-0143.

[7] S. Nakamoto, "Bitcoin: A Peer-to-Peer Electronic Cash System." Oct. 31, 2008.

[8] R. Alkhudary, "Blockchain Technology between Nakamoto and Supply Chain Management: Insights from Academia and Practice," *SSRN Journal*, 2020, doi: 10.2139/ssrn.3660342.

[9] L. Laforet and G. Bilek, "Blockchain: an inter-organisational innovation likely to transform supply chain," *Supply Chain Forum: An International Journal*, pp. 1–10, Aug. 2021, doi: 10.1080/16258312.2021.1953931.

[10] H. Treiblmaier, "The impact of the blockchain on the supply chain: a theory-based research framework and a call for action," *Supply Chain Management: An International Journal*, vol. 23, no. 6, pp. 545–559, Sep. 2018, doi: 10.1108/SCM-01-2018-0029.

[11] S. Benzidia, N. Makaoui, and N. Subramanian, "Impact of ambidexterity of blockchain technology and social factors on new product development: A supply chain and Industry 4.0 perspective," *Technological Forecasting and Social Change*, vol. 169, p. 120819, Aug. 2021, doi: 10.1016/j.techfore.2021.120819.

[12] R. Akkaoui, X. Hei, and W. Cheng, "EdgeMediChain: A Hybrid Edge Blockchain-Based Framework for Health Data Exchange," *IEEE Access*, vol. 8, pp. 113467–113486, 2020, doi: 10.1109/ACCESS.2020.3003575.

[13] V. Malamas, P. Kotzanikolaou, T. K. Dasaklis, and M. Burmester, "A Hierarchical Multi Blockchain for Fine Grained Access to Medical Data," *IEEE Access*, vol. 8, pp. 134393–134412, 2020, doi: 10.1109/ACCESS.2020.3011201.

[14] A. Yazdinejad, G. Srivastava, R. M. Parizi, A. Dehghantanha, K.-K. R. Choo, and M. Aledhari, "Decentralized Authentication of Distributed Patients in Hospital Networks Using Blockchain," *IEEE J. Biomed. Health Inform.*, vol. 24, no. 8, pp. 2146–2156, Aug. 2020, doi: 10.1109/JBHI.2020.2969648.

[15] M. Reda, D. B. Kanga, T. Fatima, and M. Azouazi, "Blockchain in health supply chain management: State of art challenges and opportunities," *Procedia Computer Science*, vol. 175, pp. 706–709, 2020, doi: 10.1016/j.procs.2020.07.104.

[16] M. H. Kassab, J. DeFranco, T. Malas, P. Laplante, giuseppe destefanis, and V. V. Graciano Neto, "Exploring Research in Blockchain for Healthcare and a Roadmap for the Future," *IEEE Trans. Emerg. Topics Comput.*, pp. 1–1, 2019, doi: 10.1109/TETC.2019.2936881.

[17] A. Farouk, A. Alahmadi, S. Ghose, and A. Mashatan, "Blockchain platform for industrial healthcare: Vision and future opportunities," *Computer Communications*, vol. 154, pp. 223–235, Mar. 2020, doi: 10.1016/j.comcom.2020.02.058.

[18] M. Niranjanamurthy, B. N. Nithya, and S. Jagannatha, "Analysis of Blockchain technology: pros, cons and SWOT," *Cluster Comput*, vol. 22, no. S6, pp. 14743–14757, Nov. 2019, doi: 10.1007/s10586-018-2387-5.

[19] N. Kshetri, "Blockchain's roles in strengthening cybersecurity and protecting privacy," *Telecommunications Policy*, vol. 41, no. 10, pp. 1027–1038, Nov. 2017, doi: 10.1016/j.telpol.2017.09.003.

[20] E. J. De Aguiar, B. S. Faiçal, B. Krishnamachari, and J. Ueyama, "A Survey of Blockchain-Based Strategies for Healthcare," *ACM Comput. Surv.*, vol. 53, no. 2, pp. 1–27, Jul. 2020, doi: 10.1145/3376915.

[21] T. McGhin, K.-K. R. Choo, C. Z. Liu, and D. He, "Blockchain in healthcare applications: Research challenges and opportunities," *Journal of Network and Computer Applications*, vol. 135, pp. 62–75, Jun. 2019, doi: 10.1016/j.jnca.2019.02.027.

[22] T. Benil and J. Jasper, "Cloud based security on outsourcing using blockchain in E-health systems," *Computer Networks*, vol. 178, p. 107344, Sep. 2020, doi: 10.1016/j.comnet.2020.107344.

[23] Y. Zhuang, L. R. Sheets, Y.-W. Chen, Z.-Y. Shae, J. J. P. Tsai, and C.-R. Shyu, "A Patient-Centric Health Information Exchange Framework Using Blockchain Technology," *IEEE J. Biomed. Health Inform.*, vol. 24, no. 8, pp. 2169–2176, Aug. 2020, doi: 10.1109/JBHI.2020.2993072.

[24] S. Shi, D. He, L. Li, N. Kumar, M. K. Khan, and K.-K. R. Choo, "Applications of blockchain in ensuring the security and privacy of electronic health record systems: A survey," *Computers & Security*, vol. 97, p. 101966, Oct. 2020, doi: 10.1016/j.cose.2020.101966.

[25] K. Peffers, T. Tuunanen, M. A. Rothenberger, and S. Chatterjee, "A Design Science Research Methodology for Information Systems Research," *Journal of Management Information Systems*, 2014, doi: DOI: 10.2753/MIS0742-1222240302.

[26] Hevner, March, Park, and Ram, "Design Science in Information Systems Research," *MIS Quarterly*, vol. 28, no. 1, p. 75, 2004, doi: 10.2307/25148625.

[27] C. Okoli, "A Guide to Conducting a Standalone Systematic Literature Review," *CAIS*, vol. 37, 2015, doi: 10.17705/1CAIS.03743.

[28] J. Webster and R. T. Watson, "Analyzing the Past to Prepare for the Future: Writing a Literature Review," *MIS Quarterly*, vol. 26, no. 2, pp. xiii–xxiii, 2002, doi: 10.2307/4132319.

[29] D. Tranfield, D. Denyer, and P. Smart, "Towards a Methodology for Developing Evidence-Informed Management

 ISBN: 978-1-61208-916-4

Knowledge by Means of Systematic Review," *British Journal of Management*, vol. 14, no. 3, pp. 207–222, Sep. 2003, doi: 10.1111/1467-8551.00375.

 ISBN: 978-1-61208-916-4

Navi Campus: Quantitative Methodology for Evaluating the User Interface of a Navigation App Using Eye Tracker and Smartphone

Jesus Zegarra Flores
Research and Innovation
Capgemini Engineering
Illkirch, France
email:jesus.zegarraflores@capgemini.com

Sabine Cornus
Faculté des Sciences du Sport
Université de Strasbourg
Strasbourg, France
email:cornus@unistra.fr

Emma Charbonnier
Research and Innovation
Capgemini Engineering
Illkirch, France
email:emma.charbonnier@capgemini.com

Laurence Rasseneur
Faculté des Sciences du Sport
Université de Strasbourg
Strasbourg, France
email:rassene@unistra.fr

***Abstract*— Navi Campus is a mobile app developed between Capgemini engineering and the Strasbourg university. Currently, navigation apps can have some limitations; for instance, the GPS coordinates of the destination do not correspond to the real destination in a university campus and they are not adapted to people having disabilities. Our version of the app, which overcomes the problems mentioned before, is in advanced stage and we want to evaluate the use of the app in a quantitative way using a mix of data (IMU, GPS and cameras) coming from the eye tracker Tobii and the app. The idea of the eye tracker is that we can detect the ocular and head movements in addition to the information from the mobile app. In the first part of the article, we will show the kinds of data we have chosen from the app and from the eye tracker to have quantitative analysis from the speed, the step frequency, and the time that a person looks at the phone. For validating our quantitative method, these firsts tests have been conducted with three people with different profiles in three different paths of similar characteristics from the Strasbourg Campus University. This methodology seems promising for analyzing the efficiency of the app establishing the relationship between the quantitative information and the behavior of the users.**

Keywords-User activity; GPS; navigation; eye tracker; mobile app.

I. INTRODUCTION

In the Strasbourg university campus in France, common mobile navigation application like Google Maps are not precise enough to locate the building's entries and to provide wheelchair-accessible paths. These are the reasons why the mobile application Navi Campus was developed [1]. This application provides outdoor navigation to help freshman students and visitors to go to a specific building on the campus from inside or outside the campus. Indeed, Navi Campus provides timetables from buses and tramways around the campus. The application is now at an advanced stage of development, and this paper introduces a quantitative method to evaluate the performances of the application.

Indeed, qualitative and quantitative studies allow to evaluate the feeling and behavior of the users [2], [3] about the interface. This evaluation is an integral part of the development of an application, in order to improve the quality of the tool and its usability. For navigation applications, the interface is even more important as misunderstanding information can lead the user to get lost. From this perspective, we conduct a first quantitative study of the interactions between the user and the Navi Campus app.

This article describes the context of the experimentation in Section 2. Section 3 proposes a methodology choosing the data for analyzing the use of the application and the behavior of the user, based on eye tracking sensor and smartphone IMU. Section 4 presents the results and Section 5 presents the conclusions and perspectives.

II. CONTEXT OF THE EXPERIMENT

A. *Environment of the study.*

In Figure 1, three paths are defined across the University Campus of Strasbourg, which is a calm environment with no cars and uncrowded.

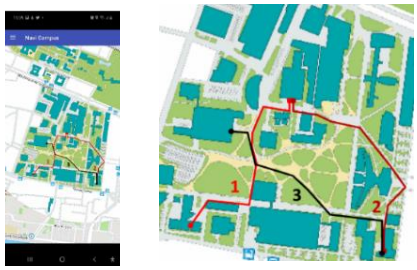


Figure 1. Three paths for the tests using Navi Campus

The first path is 400m long from "Le portique" to "SUAPS" and includes 3 sharp turns and 1 crossroad. The second path is 450m long from "SUAPS" to "Pangloss" and includes 2 sharp turn and 1 smooth turns. The third path is

 ISBN: 978-1-61208-916-4

440m long from “Pangloss” to “UFR Mathematic Info” and includes 3 sharp turns and 1 smooth turn.

B. *Subjetcs and Material*

Three subjects, who use smartphones every day, are asked to use the app to go through all three paths in the same order. The indication given was just to follow the information given by the app. Subjects have different knowledge of the campus:

Subject A knows neither the campus nor the application.

Subject B knows the campus but not the application.

Subject C knows the campus and the application.

While they are walking using the phone horizontally on the hand, the app records IMU data from the phone with a sample time of 20ms and the GPS data with a sample time of 1s. The participants also wear the Tobii Pro Glasses 2 [4] eye tracker device, which includes IMU to record head motion with a sample time of 10ms. Tobii eye tracker also provides gaze motion and video of what the user sees, which can be used to determine where the user is looking on the interface. After performing a calibration task of the eye tracker device, the subjects completed the paths and the data can be analyzed. Unfortunately, gaze data available is not exploitable as the eye tracker don’t work well outside, due to high brightness.

C. *Segmentation of the paths*

In order to sharply analyze the behavior of the users, considering the information giving by the phone, the paths are automatically segmented in turns and straight lines. To do that, the walking orientation of the subject is analyzed using the recorded GPS coordinates (after doing the path), to determine when the user has turned or not.

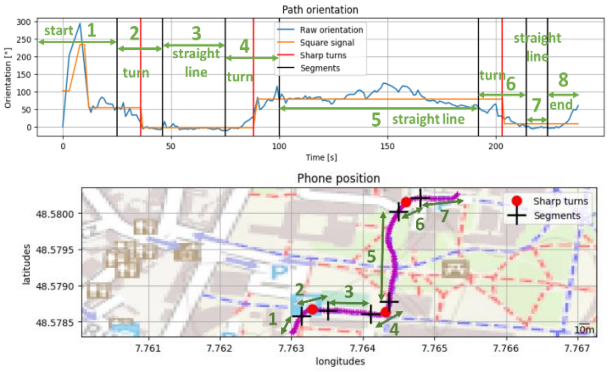


Figure 2. Example of segmented path (first path)

After approximating the orientation signal to a square signal, sharp turns can be extracted using a threshold 60°. Turn segments starts 15 meters before the turn and finishes 15m away. False turns are ignored when the user is not moving. In Figure 2, three sharp turns, 3 straight lines and start and end segments are detected (example of first path).

III. METHODOLOGY

After synchronizing the phone and the eye tracker data in post-processing, in each path segment, the speed, the steps frequency and the time when the user looks the app are identified using algorithms with Python to compare the behavior of all subjects in straight lines and turns.

A. *Speed analysis*

The user speed is calculated from the GPS coordinates, as the distance between two consecutives points calculated with the Haversine formula divided by the elapsed time. As the GPS is not very precise, some abnormous points can occur. Therefore, speeds higher than 8 km/h are ignored (Figure 3). The mean speed of each segment is then calculated and compared between segments.

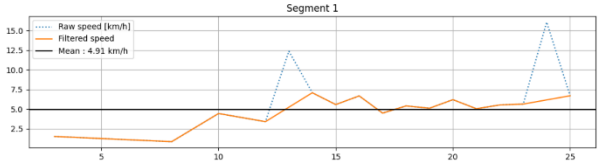


Figure 3. Example of filtered speed.

B. *Step frequency*

If the user steps frequency is low, it can be read as the user hesitating about the path to follow or an obstacle like people riding bicycles on his/her way. During the tests, the subjects hold the phone horizontally even when they are not looking at it, that is why the detection is based on the phone acceleration rather than the eye tracker acceleration, as the head motion induce too much acceleration variations.

To identify if the user is walking at a low steps frequency, the compound acceleration is calculated.

$$a_{comp} = a_x + a_y + a_z \quad (1)$$

To reduce noise, an average filter (window of 40 samples) is applied. As the sample time is not stable, no frequency filter is employed. The average signal is used to define a variable threshold for steps detection. 1.7m/s² is an empiric value determined in indoor environnment, that gives less than 10% error. When the compound acceleration is higher than the threshold, a step is potentially detected (Figure 4 (a)). If this potential step is made of at least 4 samples, it is a valid step (Figure 4 (b)). The delay between two consecutives steps allows to determine a local frequency (Figure 4 (c)).

$$freq_{local}[i] = \frac{1}{time_{step_{i+1}} - time_{step_i}} \quad (2)$$

The mean steps frequency for each segment is used to know if the user is walking at a low frequency. When the local step frequency is lower than 0.5*mean steps frequency, a low frequency event is raised.

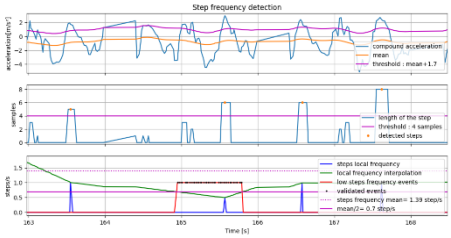


Figure 4. User low steps frequency detection . (a) Potential step detection, (b) step validation, (c) low steps frequency detection

As the accelerometer can miss some acquisitions, a confidence indicator is used to determine whether the low frequency event is due to low frequency or miss acquisions. The low frequency events due to miss acquisition are ignored.

C. Looking at the phone

When the user looks at his phone, he/she puts his/her head down and the phone appears on the eye tracker video. To detect these events, two algorithms are used.

1) Head motion

It is obsverved through the glasses' IMU. When the user looks down, the gyroscope senses a rotation along the lateral axis X and the accelerometer a translation along the 2 axis Y and Z (see Figure 5).

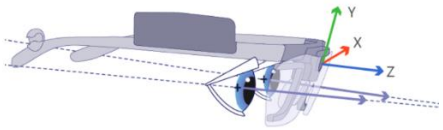


Figure 5. Glasses parameterization. [4]

To detect rotation along X, the signal is filtered with a average filter (window of 50 samples) to reduce noise. If the average variance (window of 50 samples) is higher than the variance of the whole segment, a rotation of the head is detected (see Figure 6).

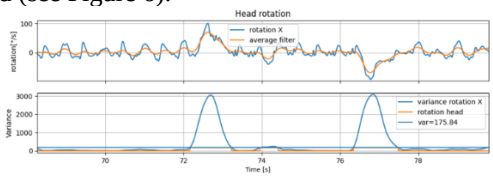


Figure 6. Detection of the head rotation

For the translation, the compound acceleration is defined as $a_{comp} = a_y + a_z$. The noise is reduced with an average filter (window of 200 samples). This mean is coupled to a square signal shown Figure 7 (top).

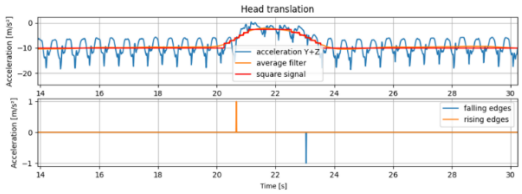


Figure 7. Detection of the head translation

When a rising edge of the square signal is followed by a falling edge, and both occur at the same time as a rotation event, an head down event is identified. For example, Figure 8 shows 11 head down detection in the first path of subject C.

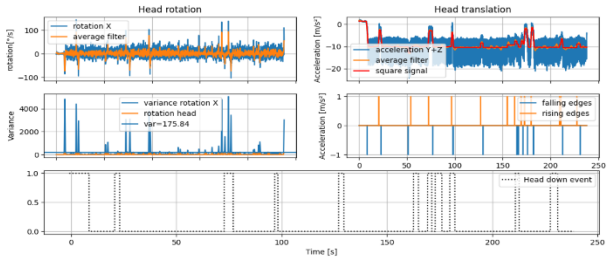


Figure 8. Detection of the head down event

2) Phone detection

When the user looks down, it does not always mean that he/she is looking at the application. A validation of the phone is made using the video. Each 400 milliseconds frame of the eye tracker is analyzed by a neural network mask RCNN [5] in mode transfert learning using the dataset coco to determine whether or not the user is looking at his phone.

As the conditions of the study were sunny, the brightness disturbs the detection ; in some cases, the phone is not detected or identified as other objects (knife, tie, skateboard, snowboard). Some exemples are presented in Figure 9.

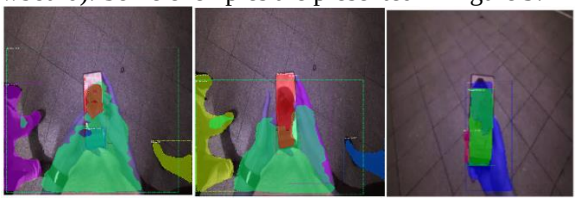


Figure 9. Exemples of identification as a cellphone (left), a tie or knife (center) and a snowboard (right)

The missidentifications can occurs quite often for some users. Indeed, as the subject A did all his paths on bright conditions, the phone is highly misstaken as a tie or a knife.

TABLE I. APPEARANCE PERCENTAGES OF EACH CLASS CONFUSED WITH THE CELL PHONE CLASS FOR ALL PATHS

Percentage appearance	Class detected				
	Cell phone	***Tie***	***Knife***	***Snowboard***	***Skateboard***
Subject A	43,976	25,000	23,494	3,916	3,614
Subject B	55,631	17,342	0,450	18,468	8,108
Subject C	75,824	0,000	3,297	8,791	12,088

To avoid losing too much data by accepting only the cell phone class, we define a superclass phone, regrouping the class *cell phone*, *tie*, *knife*, *snowboard* and *skateboard*.

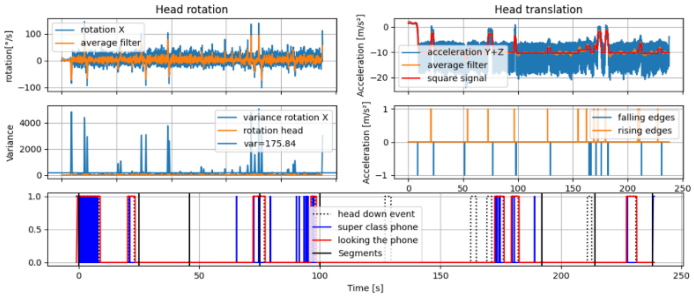


Figure 10. Detection of the user looking down to his phone

When the user looks down and an object of the super class is detected, the user is looking at his/her phone (see Figure 10).

IV. RESULTS

As all the subjects arrive at their destinations with no major problems, the application is globally efficient. To analyze the subjects' behavior, the previous algorithms are applied to the data, and the results are summarized on table 2 for the first path. The duration of the events (low steps frequency and time looking at the phone) is formulated as percentage to the segment duration.

 ISBN: 978-1-61208-916-4

TABLE II. INFORMATION FOR THE FIRST PATH IN EVERY SEGMENT FROM THE THREE SUBJECTS

Path 1	Subject	Segment duration [s]	Mean speed [km/h]	Mean step frequency [step/s]	Low step frequency			Looking the phone	
					Ratio [%]	Mean time duration [s]	Mean slow step frequency [step/s]	Mean time duration [s]	ratio [%]
1: Start	A	29.0	4.89	0.83	32.24	9.35	0.06	6.12	63.36
	B	35.49	2.59	0.11	90.88	32.26	0.02	17.08	96.24
	C	25.0	**4.91**	1.04	37.28	9.32	0.07	2.36	**9.46**
2 : Turn	A	25.0	4.88	1.12	0	0	N.A.	0	**0**
	B	24.0	5.3	0.63	30.04	7.21	0.03	1.91	15.95
	C	21.0	**5.95**	1.43	0	0	N.A.	0	**0**
3 : Straight line	A	37.0	4.9	1.43	0	0	N.A.	1.96	10.58
	B	43.0	5.57	1.35	6.51	2.8	0.54	2.9	6.74
	C	29.0	**5.6**	1.28	4.14	1.2	0.04	2.3	7.92
4 : Turn	A	32.0	4.08	1.44	0	0	N.A.	2.84	17.78
	B	27.0	4.92	1.2	0	0	N.A.	5.99	22.17
	C	25.0	**5.1**	1.01	0	0	N.A.	1.5	6
5 : Straight line	A	114.0	4.68	1.36	0	0	N.A.	2.72	19.09
	B	91.0	5.65	1.34	1.27	1.16	0.19	4.81	15.85
	C	92	**5.78**	1.39	0.91	0.84	0.59	2.79	6.06
6 : Turn	A	28.0	4.55	1.25	0	0	N.A.	**2.61**	**27.91**
	B	23.0	5.03	0.78	2.78	0.64	0.36	**3.53**	**30.66**
	C	22.0	**5.23**	0.89	0	0	N.A.	0	**0**
7 : End	A	48.0	4.32	0.83	9.19	4.41	0.25	2.74	22.86
	B	28.0	5.22	1.04	4.93	1.38	0.33	0	0
	C	24.0	**5.83**	0.8	4	0.96	0.02	3.5	**14.58**

For all paths, the user who knew best the environment and the app (C) is generally the fastest to get to the destination and the one who looks in average the least the application. On the segment 6, subjects A and B have almost the same percentage to look at theirs phones, but the mean time is higher for the subject B. An explanation can be that the subject A looks often at the phone and quickly, as subject B looks to the phone but more time.

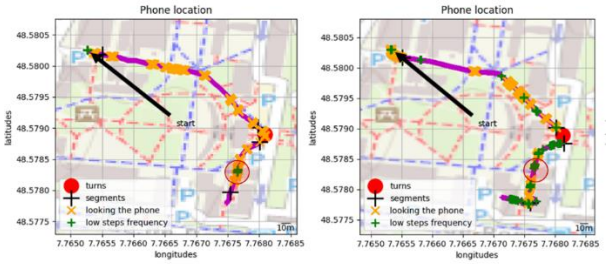


Figure 11. The person slows down (+ in green) and looking the phone detection (x orange) when pop up notification appears

In the second path, the notification of the destination appears suddenly in front of another building's entry, which disturbs the users (Figure 11). Indeed, subjects A and B slow down and look at the phone to be sure if it is the good entry. The third path does not show any abnormality.

V. CONCLUSIONS AND FUTURE WORK

The proposed evaluating method seems to be robust as it works with several users' behaviors on different paths and segments. The notification, which appears suddenly, can disturb the user in ambiguous situations. Moreover, on the first path, the user B who knew the environment would not have taken necessarily the same path as shown by the app.

Using these first results, we have started to have first interpretation of the behaviors of the subjects using the app.

The phone detection could be improved by training a specific neural network for this issue. The eye tracker camera sensors for gaze and video are sensitive to brightness, it is recommended to test at the end of the day.

We will test our methodology in more complex environments to validate it. Moreover, some experiments will be done with disabled people to choose the appropriate sensors to analyze the user interface.

Finally, a qualitative questionnaire could be associated with this quantitative study to better understand the behavior and decisions during navigation and to have a feedback after the navigation experience.

ACKNOWLEDGMENT

We would like to thank the entire Research medic@ team from Capgemini Engineering, particularly Julien Balbierer and Eric Bournez.

REFERENCES

[1] J. Zegarra Flores, F. Pereme, J.-P. Radoux and L. Rasseneur, "Navi Campus : an enhanced GPS navigation app for University Campuses", 12th ITS European Congress, abstract TP0815, Strasbourg, June 2017, pp. 37

[2] R. Fryers, T. Holzer Saad and J. Dinsmore, "Report Defining The Needs of Stakeholders For a Wayfinding Platform For Individuals With Intellectual Disabilities and Their Carers," Trinity College Dublin, Trinity College Dublin, March 2018, pp. 1-29.

[3] I. Maly, J. Balata, O. Krejcir, E. Fuzessery and Z. Mikovec, "Qualitative measures for evaluation of navigation applications for visually impaired", 6th IEEE International Conference on Cognitive Infocommunications (CogInfoCom), 2015, pp. 223-228

[4] Tobii AB, User's manual Tobii Pro Lab, Version 1.130, 12/2019., pp 82, http://www.vinis.co.kr/TPL_manual.pdf

[5] K. He, G. Gkioxari, P. Dollàr and R. B. Girshick, "Mask R-CNN", *IEEE International conference on computer vision ICCV*, pp. 2980-2988, 2017.

 ISBN: 978-1-61208-916-4

Automatic Recognition of Continuous Signing of Brazilian Sign Language for Medical Interview

Robson Silva de Souza
Dept. of Computer Engineering and Industrial Automation
School of Electrical and Computer Engineering
Campinas, SP, Brazil
robsonnddesouza@gmail.com

José Mario De Martino
Dept. of Computer Engineering and Industrial Automation
School of Electrical and Computer Engineering
Campinas, SP, Brazil
martino@unicamp.br

Janice Gonçalves Temoteo Marques
Dept. of Human Development and Rehabilitation
Faculty of Medical Sciences, University of Campinas
Campinas, SP, Brazil
janicetm@unicamp.br

Ivani Rodrigues Silva
Dept. of Human Development and Rehabilitation
Faculty of Medical Sciences, University of Campinas
Campinas, SP, Brazil
ivanirs@unicamp.br

***Abstract*—In this article, we present an automatic image recognition approach for assisting the communication between deaf patients, speakers of the Brazilian Sign Language (Libras), and hearing physicians. The aim of the approach is to help the interaction and exchange of information during medical interviews. Its scope is the automatic recognition of the continuous signing of Libras through the analysis of traditional video and depth data (RGB-D data). Recognition is performed by a cascade of two neural networks. The first, a convolutional neural network, encodes the visual input and extracts relevant features. The second, a recurrent neural network, learns the mapping of the extracted features into Brazilian Portuguese words. To train the recurrent network with videos of different lengths and word sequences, we use the Connectionist Temporal Classification approach. Experiments using a dataset of 280 videos encompassing 56 sentences composed of 67 different signs results in an accuracy of round 91%.**

***Index Terms*—Libras, Sign language recognition, Continuous signing, long short term memory, connectionist temporal classification**

I. Introduction

Anamnesis and clinical examination are the standard procedures of physicians to diagnose diseases and health problems of their patients. Anamnesis is a process of interviewing the patient to collect information about his/her current health complaints and medical history. The precise disclosure, correct understanding, and assessment of this information are preconditions for an effective diagnosis and the identification of the appropriate therapy. However, the effectiveness of the medical interview is jeopardized if the physician and the patient do not have a common language for communication. That is usually the case when we consider a deaf patient who has sign language as his/her first language and does not master the written language of the physician who, by his/her side, does not understand sign language. A common solution to overcome this problem is to have a sign language interpreter assisting the deaf patient during the interview. Besides the operational difficulties of organizing an interpreter, another important drawback is the uncomfortable situation created by the introduction of a third party in the medical interview. During a medical interview, the patient should feel comfortable enough to share very personal and sensitive information, providing any and all relevant information to help the doctor to make a correct diagnosis. A solution to overcome this potential breach of patient-doctor confidentiality is to provide a robust computer-based solution to support the communication between physicians and deaf patients. Although the interaction between doctor and patient is a two-way process, in this article, we focus only on the issue of automatic recognition of continuous signing based on computer-based recognition of video imagery. Our study deals specifically with the Brazilian Sign Language. However, the findings can be extended and applied to other sign languages.

Sign languages convey information by the movement of the hands, body, and face. They are perceived by vision. There is not a single, universal sign language used worldwide by deaf people. Each country has its own sign language [1]. The sign language of a country is independent of its oral language. For example, Deaf Americans speak the American Sign Language (ASL), the Deaf in the UK use the British Sign Language (BSL), and Deaf Australians speak the Australian Sign Language (Auslan). Deaf Brazilians use the Brazilian Sign Language (Libras).

There is an increasing research interest in automatic sign language recognition in recent years. Automatic sign language recognition applies computer vision combined with machine learning techniques to analyze and translate, into a written form, videos with sign language content.

The development of robust automatic sign language recognition systems is challenging. Several techniques have been proposed for automatic sign language recognition for a variety of sign languages, including the Brazilian Sign Language (Libras). Most efforts, however, have been limited to the study of isolated sign recognition, postures representative of cardinal

 ISBN: 978-1-61208-916-4

numbers (0 to 10), and the manual alphabet or fingerspelling. Research on continuous signing recognition is still rare.

In this article, we present a method for automatic continuous sign language recognition of Libras during medical interviews. Applying the method, we implement an approach based on Deep Learning that is capable of finding and using extracted data from signing from full-frame sequences. Therefore, it aligns sequences of video frames displaying Libras content to sequence glosses. A gloss is a word, in our case a Portuguese word, that is consistently used to label a sign within the corpus, regardless of the meaning of that sign in a particular context or whether it has been systematically modified in some way [2]. As pointed out in [3], glosses are a convenient way to write down the meaning of a sign, as they use another language to represent the signs.

The main contributions of this article are:

- The construction of a robust and representative dataset, composed of RGB information and depth of signage in Libras in order to contribute to the advancement of the research in this area.
- Execution of a Depth-Wise Separable Convolutional Network (DWSCN) based architecture, as feature extractor preprocessor. Insofar as we know, we are the first to employ this type of architecture in continuous sign language recognition systems.
- The development of a new architecture of sequential learning, based on recurrent neural networks and Connectionist Temporal Classification (CTC), which learn to find and store relevant data in its memory cells from the full-frame sequences, without importing in its subsystems structures that process image patches.

The remainder of the paper is organized as follows: Section II contains a review of relevant related work. Section III presents our approach. Section IV describes the experiments performed, and Section V presents the conclusions.

II. Related Work

The recognition of continuous signing is a far more complex task than the recognition of isolated signs, requiring more sophisticated methods to deal with the dynamics of production and the transition between signs. On the other hand, continuous signing recognition systems are more appropriate for real-world scenarios of interpersonal communication. However, it is observed that there is still little research that seeks to solve this problem. In the following paragraphs, we present approaches aimed at recognizing continuous signing based on computer vision.

Research using deep learning models has increased considerably in recent years. The work of [4] proposes an approach that breaks down the problem of recognizing signs into a series of expert systems called subunits. Each subunit consists of three layers of neural networks; Convolutional Neural Network (CNN) for extraction of spatial features, Bidirectional Long Short-Term Memory (BLSTM) [5], an extension of LSTM [6] that temporarily models the features and a loss layer based on the CTC. A recent work, [7], also uses CNN and LSTM but encapsulated it in an Hidden Markov Model (HMM) model following the hybrid approach used in his previous work, this time exploring sequential parallelism to learn sign language, mouth shapes, and hand shape classifiers.

The works [8]–[12] use CNNs as feature extractors, a 3D CNN model, or a 3D residual convolutional network (3D-ResNet). For modeling and sequential learning, they use dilated convolutional networks or RNNs such as LSTM, Gated Recurrent Unit (GRU) [13] and their variants in combination with the CTC algorithm. Among these approaches, [14] is the one that achieved the best performance in the RWTH-PHOENIX-Weather dataset and also in a set of images captured by the Kinect called CSL-25K, which covers 100 daily life sentences expressed in Chinese Sign Language (CSL).

In our proposal, we also use recurrent neural networks with CTC, but differently from the other approaches, we apply depth-wise separable convolutional network that contains far fewer parameters and is computationally cheaper than the state-of-the-art convolutional neural networks, as for example VGG16 [15], ResNet50 [16] and InceptionV3 [17].

III. Method

In this section, we present the protocol used to build a sign language dataset in the context of a medical interview and our approach for recognizing continuous signing in Libras.

A. Dataset construction

The existence of a dataset composed of Libras sentences related to medical interviews is fundamental to develop and test our approach. No publicly available image databases of continuous signing in Libras have been found.

Through the study of existing datasets of other sign languages [18] and with the intent of meeting our objectives, we developed specifications to be followed for the construction of our dataset. The proposal is to develop a robust dataset that simulates the internal environment of a clinic with artificial lighting, in which the deaf volunteer or interpreter performs the sign naturally.

Fig. 1 shows the execution flow. Thereafter, each module will be described in details.

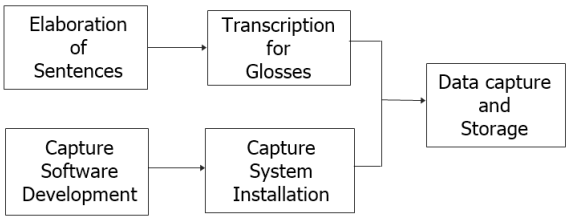


Fig. 1. Execution flow for the construction of the dataset

Sentences elaboration. It comprises the elaboration of sentences in the Portuguese language related to the answers of a patient in the context of a medical consultation (general practitioner). The sentences are established through the study of signs and manifested individual symptoms accordingly to the anamnesis medical procedure described in [19] and [20].

Transcription of the sentences in Portuguese to glosses. It is created through the assistance of a fluent sign language specialist. The right columns on Tables I and II present the transcriptions of the sentences from the previous stage.

TABLE I
EXAMPLES OF SENTENCES DEVISED IN PORTUGUESE LANGUAGE AND ITS TRANSCRIPTIONS TO GLOSSES.

#	Target	Prediction
1	Eu tenho febre	EU FEBRE
2	Eu estou fraco	EU FRACO
3	Eu estou com diarréia	EU TER DIARRÉIA
4	Meu braço esquerdo dói	MEU BRAÇO-ESQUERDO DOR
5	Minha urina está marrom	MEU XIXI COR MARROM

TABLE II
EXAMPLES OF SENTENCES DEVISED IN PORTUGUESE LANGUAGE AND ITS TRANSCRIPTIONS TO GLOSSES - VERSION IN ENGLISH

#	Target	Prediction
1	I have fever	ME FEVER
2	I am weak	ME WEAK
3	I have diarrhea	ME HAVE DIARRHEA
4	My left arm hurts	MY LEFT-ARM PAIN
5	My urine is brown	MY PEE BROWN COLOR

Capture device and development of the capture software. The data recording is made through the Kinect device v2 for Windows. The capture application is developed using Kinect's own software development kit (SDK). This application captures and stores RGB images, depth images, and mapped images (RGB images mapped on the depth images), in which all pixels not belonging to the signer are converted to black.

Installation of the capture system. The Libras signing recordings executed by the volunteer are made in a laboratory, with artificial illuminations and homogeneous scene background.

Capture and data storage. A Libras interpreter teacher, member of the research team helps with the video acquisition. During the signing the images are captured and stored on the computer.

B. Our approach

The approach that recognizes continuous Libras signing includes a CNN-based model for features extraction and an RNN architecture for learning the spatial-temporal dependencies that exist between the sentence signs. To solve the alignment problem between the probability sequences in the RNN outputs with the sequences of glosses, we used CTC.

Fig. 2 presents a general view of our approach composed of three main models. The first comprises spatial modeling, while the others encompass sequential learning and a CTC loss layer to decode categorical probabilities in sequences of glosses.

Features extraction. DWSCN is used for representations of spatial features of the frame sequences. The pre-trained MobileNetV1 [21] operational model is among the models

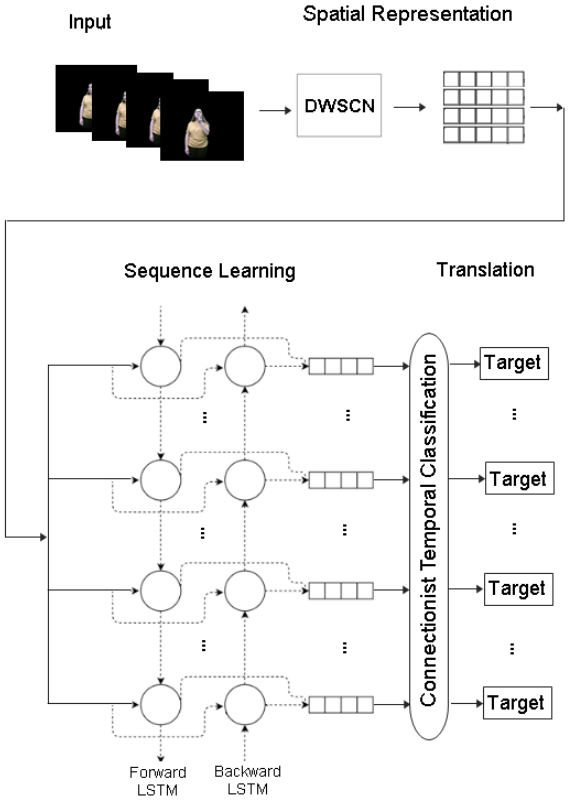


Fig. 2. Overview of our continuous sign language recognition approach

based on the DWSCN. The use of pre-trained models enables developing efficient models in situations of limited data availability, in addition to reducing processing time [22].

MobileNetV1 was pre-trained on ImageNet [23] and has characteristics of having reduced size (17MB) and reduced number of parameters (4,2 million) when compared to other state-of-the-art models.

To use MobileNet as a feature extractor preprocessor, the softmax classification layer (SM) and the completely connected layer (FC) have been removed, keeping all the depthwise separable convolution blocks and the Average Pooling layer.

All dataset images are processed by the resulting model. As the last layer has 1024 nodes, each image will be represented as a 1024 value vector. Each video sample results in a three-dimensional array of dimensions equal to 1 x number of frames x 1024 features. Since the number of frames are different between the videos, the padding in each array has been performed to allow the concatenating of all feature arrays. The data is used as the entry to train our model based on recurrent neural networks.

The glosses are coded in categorical variables and together with the feature arrays are used as input to train our model based on recurring neural networks. This is a weakly supervised learning problem, that is, the gloss sequences are available but not its time limits.

Sequential learning. Our approach uses BLSTM to model

 ISBN: 978-1-61208-916-4

the correspondences between the input sequences and output glosses. This architecture is capable of storing data for long periods of time and try to avoid the explosion of the gradient, a common problem of the Vanilla neural networks.

To implement a BLSTM network, it takes two parallels layers of LSTM cells, backward LSTM and forward LSTM, each of them being responsible for processing the information in the direction of time. The final hidden layer is given by the concatenation of the two networks.

The memory neurons of an LSTM are called cells. Fig. 3 presents the structure of a BLSTM network and highlights one single memory cell. The cells are capable of storing data in the course of a sequence through units called gates. According to [24], these units calculate the weights that connect them to avoid the gradient degradation through parameterized or manually chosen values.

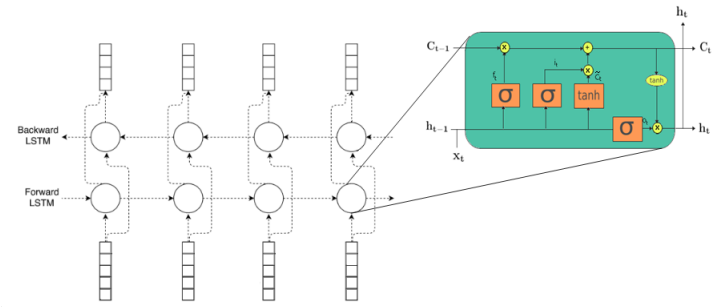


Fig. 3. BLSTM network structure, highlight to a single memory cell

A softmax activation function on a fully connected layer is used in the network output and is applied to each time frame.

Connectionist Temporal Classification. In the BLSTM training phase, CTC is used to calculate the cost value. During prediction, it decodes the probability matrices of the softmax function in gloss sequences.

To allow the CTC algorithm to decode the target sequence, one more unit is introduced to the total number of labels in the softmax output layer. This unit refers to a token named blank, that models the transitions between different labels.

Let us consider the mapping of the input frames sequence $X = [x_1, x_2, ..., x_T]$, for the sequences of output words $Y = [y_1, y_2, ..., y_T]$. The CTC cost function for a pair (X, Y) has the conditional probability $p(Y/X)$ equal to the sum of all the valid paths $A \in A_{XY}$, calculating the probability $p_t(a_t|X)$ to a single step-by-step alignment following (1).

$$p(Y/X) = \sum_{A \in A_{XY}} \prod_{a}^{b} p_t(a_t|X) \tag{1}$$

For a training set M, the model parameters are tuned to minimize the negative log-likelihood. That way, the CTC objective function is given by (2).

$$Loss_{CTC} = \sum_{(X,Y) \in M} -\log p(Y/X) \tag{2}$$

To calculate the CTC loss efficiently, the Forward-Backward algorithm given in [25] is used.

IV. Experiments

This section reports on the experiments performed and the performance of our architecture in continuous Libras signing recognition.

A. Dataset

In order to develop and test our approach, 280 sentences signed by a professional interpreter were captured, corresponding to 5 repetitions of 56 sentences. 42663 frames are obtained at a rate of 30 fps. The number of glosses is equal to 67. The number of glosses per sentence varies from 2 to 6 and the number of frames per sentence varies from 124 to 277.

B. Evaluation Metrics

The Word Error Rate (WER) is the metric widely [8], [9], [26], [27], [12], [10], [11], [14], [4], [28], [29], [30] used in continuous sign language recognition work and, therefore, will be the metric used in this paper. The WER is given by (3).

$$WER = \frac{I + D + S}{N} \tag{3}$$

Where I is the number of errors entered, D is the number of deletion errors, S is the number of substitution errors, and N is the total number of glosses in the reference sentence.

The accuracy is given by (4).

$$acc = 1 - WER \tag{4}$$

C. Training and Evaluation

We performed experiments on an Nvidia RTX 2080Ti, and the model is implemented in the Keras framework [31], using tensorflow [32] as a backend. In our experiments, we used 80 percent of the data (224 sentences) for the training set and 20 percent for the test set (56 sentences).

The simulations performed processed the images mapped. Initially, these images were resized to 224 X 224 pixels, dimensions expected by the MobileNetV1 network.

After spatial modeling with our structure based on DWSCN, the resulting feature matrix has a dimension equal to the number of samples x time steps x features.

According [33], training small datasets has some challenges, as the network effectively memorizes the training dataset. The author recommends that adding noise is an approach to improve the generalization error and to enhance the structure of the mapping problem during learning. Thus, we applied Gaussian noise, at the entrance of the BLSTM network, with a standard deviation of 0.5 during the training phase.

The training of our BLSTM architecture is performed by implementing the backpropagation algorithm through time, [34]. The initialization for the recurrent weights matrix is the orthogonal [35], for non-recurring weights the glorot uniform [36] and the vector bias is initialized with zeros. The optimizer used is the Root Mean Square Propagation (RMSprop) [37] with a learning rate of 0.01, a discounting factor of 0.9, a

 ISBN: 978-1-61208-916-4

momentum of zero, (default values in the framework), and a batch size of 82. Then, we use the CTC beam decoder described in the work of [38] to decode sentences with beam width 10.

For the aforementioned configurations, dozens of experiments were carried out using different network topologies, with a maximum of 4 layers (1 to 3 recurring layers and a completely connected layer) and the number of neurons equal to powers of 2 in the range of 2 to 512. The last layer is fixed with 68 neurons (one for each vocabulary label plus the blank label). Given the stochastic nature of the algorithms used, repetitions of the tests are performed in order to determine the most promising models.

In order to detect overfitting and determine the most promising models, a validation set is adopted, based on the training set, consisting of 60 sentences. During the training, at the end of each epoch, the value of the loss CTC is calculated in the validation set, and the best model in each training is determined according to the lowest value of the loss in that set.

Also, to identify and soften the effect of overfitting, we used the method of regularization called dropout, presented in [39]. Dropout values equal to 0.5 were applied for both recurrent and non-recurrent connections.

Among the best models that fit the data, the simplest model, that is, with the least hyperparameters, is considered the most plausible to be used in the test set.

D. Results

Our best result is achieved by configuring two recurrent layers with 32 and 64 neurons, respectively. At the end of 30000 epochs, it was determined that the best model corresponds to epoch 21422. The values of the initial weights and the settings referring to that model are saved and stored for reproducibility, as well as for use in the unseen data set during the training.

Of the 56 sentences in the test set, 11 obtained some kind of error in the model prediction. The average WER was 8.92% and therefore, an accuracy of 91.07%. In Table III we can observe some errors found, comparing the results of the model with the ground-truth sentences. Bold words are associated with errors in prediction. Table IV presents the equivalent results in English.

Therefore, the errors found were: 13 substitutions, 2 insertions, and no deletions. Low values in relation to the total amount of glosses existing in the dataset demonstrating the effectiveness of our architecture.

TABLE III
SENTENCES WITH PREDICTION ERRORS

#	Target	Prediction
1	COMEÇAR ANTEONTEM	COMEÇAR **ONTEM**
2	COMEÇAR QUINTA-FEIRA PASSADA	COMEÇAR **TERÇA-FEIRA** PASSADA
3	COMEÇAR SEGUNDA-FEIRA PASSADA	COMEÇAR **QUARTA-FEIRA** PASSADA
4	COMEÇAR TERÇA-FEIRA PASSADA	COMEÇAR **QUINTA-FEIRA** PASSADA
5	MAU-HÁLITO FEDOR TER	MAU-HÁLITO FEDOR **VERMELHO** TER
6	MEU DENTE DOR	MEU **COSTAS TER**
7	MEU NARIZ DOR	MEU **OLHO-ESQUERDO INCHADO**
8	OLHO-DIREITO APONTAR VERMELHO TER	OLHO-DIREITO APONTAR VERMELHO **SABOR NÃO-TER**
9	MEU OLHO-DIREITO DOR	MEU OLHO-DIREITO **INCHADO**

TABLE IV
SENTENCES WITH PREDICTION ERRORS - VERSION IN ENGLISH

#	Target	Prediction
1	START BEFORE-YESTERDAY	START **YESTERDAY**
2	START THURSDAY PAST	START **TUESDAY** PAST
3	START MONDAY PAST	START **WEDNESDAY** PAST
4	START TUESDAY PAST	START **THURSDAY** PAST
5	BAD-BREATH BAD-SMELL HAVE	BAD-BREATH BAD-SMELL **RED** HAVE
6	MY TOOTH PAIN	MY **BACK HAVE**
7	MY NOSE PAIN	MY **LEFT-EYE SWOLLEN**
8	RIGHT-EYE POINT RED HAVE	RIGHT-EYE POINT RED **FLAVOR DO-NOT-HAVE**
9	MY RIGHT-EYE PAIN	MY RIGHT-EYE **SWOLLEN**

V. CONCLUSIONS

In this article, we presented an approach for recognition of continuous signing of Libras. This approach receives sequences of images of a person communicating in Libras and translates signs to the Portuguese language. The efficacy of our proposed methods was proven by state-of-the-art results.

In general, when compared to other approaches in the literature, our approach demonstrates a series of advantages:

i) It does not depend on the extraction of manual features, specifically designed for a domain and laboriously calculated from the geometry of the hands and arms.

ii) It takes into account characteristics related to non-manual expressions, such as movements of the face, eyes, head, and torso, instead of using only continuous sequences of the hands.

iii) Contrary to other studies' continuous signing recognition, which performs the feature extraction process in video segments related to isolated signs, our spatial representation module is processed on the entire video. Our choice is due to the fact that video representation based on fixed-length signs can compromise the continuous recognition of signing in real situations since the same sign varies in length in a video, even when performed by the same person in different situations

iv) Our spatial modeling, which is based on depthwise separable convolutions, reduces the latency and favors the development of real-time sign recognition because of the accuracy and the number of parameters and demanded calculations. This is a great advantage when compared to other convolutional neural networks.

 ISBN: 978-1-61208-916-4

v) Our architecture based in BLSTM with CTC learns to find and store information relevant memory cells from the data channels included in full-frame sequences. This is done without injecting subsystems in its structure that process image patches. Consequently, our approach presents a greater capacity for temporal learning compared to studies that import extra data in its system to ease the learning.

Our approach demonstrates the potential to be applied in signing recognition on heterogeneous backgrounds due to the use of Kinect, which performs the segmentation of the individual while capturing the depth and color of images. In our upcoming work, we intend to include more signage and diversify the recording scenarios of our dataset images, as well as increase the vocabulary in order to maximize the robustness of our recognition approach.

ACKNOWLEDGMENT

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nivel Superior - Brasil (CAPES) - Finance Code 001.

REFERENCES

[1] N. Timmermans *et al.*, *The status of sign languages in Europe*. Council of Europe, 2005.

[2] T. Johnston, "From archive to corpus: transcription and annotation in the creation of signed language corpora," in *Proceedings of the 22nd Pacific Asian Conference on Language, Information, and Computation*, pp. 16–29, 2008.

[3] A. Baker, B. van den Bogaerde, R. Pfau, and T. Schermer, *The linguistics of sign languages: An introduction*. John Benjamins Publishing Company, 2016.

[4] N. C. Camgoz, S. Hadfield, O. Koller, and R. Bowden, "Subunets: End-to-end hand shape and continuous sign language recognition," in *2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 3075–3084, IEEE, 2017.

[5] A. Graves and J. Schmidhuber, "Framewise phoneme classification with bidirectional lstm and other neural network architectures," *Neural networks*, vol. 18, no. 5-6, pp. 602–610, 2005.

[6] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[7] O. Koller, C. Camgoz, H. Ney, and R. Bowden, "Weakly supervised learning with multi-stream cnn-lstm-hmms to discover sequential parallelism in sign language videos," *IEEE transactions on pattern analysis and machine intelligence*, 2019.

[8] J. Pu, W. Zhou, and H. Li, "Dilated convolutional network with iterative optimization for continuous sign language recognition.," in *IJCAI*, pp. 885–891, 2018.

[9] J. Pu, W. Zhou, and H. Li, "Iterative alignment network for continuous sign language recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4165–4174, 2019.

[10] R. Cui, H. Liu, and C. Zhang, "Recurrent convolutional neural networks for continuous sign language recognition by staged optimization," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7361–7369, 2017.

[11] R. Cui, H. Liu, and C. Zhang, "A deep neural framework for continuous sign language recognition by iterative training," *IEEE Transactions on Multimedia*, 2019.

[12] H. Zhou, W. Zhou, and H. Li, "Dynamic pseudo label decoding for continuous sign language recognition," in *2019 IEEE International Conference on Multimedia and Expo (ICME)*, pp. 1282–1287, IEEE, 2019.

[13] K. Cho, B. Van Merriënboer, D. Bahdanau, and Y. Bengio, "On the properties of neural machine translation: Encoder-decoder approaches," *arXiv preprint arXiv:1409.1259*, 2014.

[14] H. Zhou, W. Zhou, Y. Zhou, and H. Li, "Spatial-temporal multi-cue network for continuous sign language recognition," *arXiv preprint arXiv:2002.03187*, 2020.

[15] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[16] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

[17] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision. 2015," *arXiv preprint arXiv:1512.00567*, 2015.

[18] N. B. Ibrahim, H. H. Zayed, and M. M. Selim, "Advances, challenges and opportunities in continuous sign language recognition," *Journal of Engineering and Applied Sciences*, vol. 15, no. 5, pp. 1205–1227, 2020.

[19] F. Veiga and A. B. Souza, *Physical Exam Manual*. Elsevier Brasil, 2019.

[20] M. H. Swartz, *medical semiology treatise*. Elsevier Brasil, 2015.

[21] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," *arXiv preprint arXiv:1704.04861*, 2017.

[22] J. Brownlee, *Deep Learning for Computer Vision: Image Classification, Object Detection, and Face Recognition in Python*. Machine Learning Mastery, 2019.

[23] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*, pp. 248–255, Ieee, 2009.

[24] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. The MIT Press, 2016.

[25] A. Graves, S. Fernández, F. Gomez, and J. Schmidhuber, "Connectionist temporal classification: labelling unsegmented sequence data with recurrent neural networks," in *Proceedings of the 23rd international conference on Machine learning*, pp. 369–376, ACM, 2006.

[26] D. Guo, W. Zhou, H. Li, and M. Wang, "Hierarchical lstm for sign language translation," in *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[27] J. Huang, W. Zhou, Q. Zhang, H. Li, and W. Li, "Video-based sign language recognition without temporal segmentation," in *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[28] O. Koller, R. Bowden, and H. Ney, "Automatic alignment of hamnosys subunits for continuous sign language recognition," *LREC 2016 Proceedings*, pp. 121–128, 2016.

[29] O. Koller, H. Ney, and R. Bowden, "Deep hand: How to train a cnn on 1 million hand images when your data is continuous and weakly labelled," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3793–3802, 2016.

[30] O. Koller, O. Zargaran, H. Ney, and R. Bowden, "Deep sign: Hybrid cnn-hmm for continuous sign language recognition," in *Proceedings of the British Machine Vision Conference 2016*, 2016.

[31] F. Chollet *et al.*, "Keras." https://keras.io, last accessed on 02/08/21, 2015.

[32] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, *et al.*, "Tensorflow: Large-scale machine learning on heterogeneous distributed systems," *arXiv preprint arXiv:1603.04467*, 2016.

[33] J. Brownlee, *Better Deep Learning: Train Faster, Reduce Overfitting, and Make Better Predictions*. Machine Learning Mastery, 2018.

[34] P. J. Werbos, "Backpropagation through time: what it does and how to do it," *Proceedings of the IEEE*, vol. 78, no. 10, pp. 1550–1560, 1990.

[35] A. M. Saxe, J. L. McClelland, and S. Ganguli, "Exact solutions to the nonlinear dynamics of learning in deep linear neural networks," *arXiv preprint arXiv:1312.6120*, 2013.

[36] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, pp. 249–256, 2010.

[37] G. Hinton, N. Srivastava, and K. Swersky, "Neural networks for machine learning lecture 6a overview of mini-batch gradient descent," *Cited on*, vol. 14, p. 8, 2012.

[38] A. Hannun, C. Case, J. Casper, B. Catanzaro, G. Diamos, E. Elsen, R. Prenger, S. Satheesh, S. Sengupta, A. Coates, *et al.*, "Deep speech: Scaling up end-to-end speech recognition," *arXiv preprint arXiv:1412.5567*, 2014.

[39] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov, "Improving neural networks by preventing co-adaptation of feature detectors," *arXiv preprint arXiv:1207.0580*, 2012.

 ISBN: 978-1-61208-916-4