A Hybrid Machine Learning Approach for Enhanced Skin Cancer Diagnosis Using Convolutional Neural Networks, Support Vector Machines, and Gradient Boosting

Fazila Faizal Patel
School of Engineering, Technology and Design
Canterbury Christ Church University
Canterbury, United Kingdom
E-mail: f.patel149@canterbury.ac.uk

Amina Souag
School of Engineering, Technology and Design
Canterbury Christ Church University
Canterbury, United Kingdom
E-mail: amina.souag@canterbury.ac.uk

Adedayo Olowolayemo
School of Engineering, Technology and Design
Canterbury Christ Church University
Canterbury, United Kingdom
E-mail: a.olowolayemo502@canterbury.ac.uk

Abstract*— This study investigates the effectiveness of a hybrid machine learning model for skin cancer diagnosis, integrating Convolutional Neural Networks, Support Vector Machines, and Gradient Boosting algorithms. By combining the strengths of each technique, the model seeks to improve diagnostic accuracy and reliability in clinical settings, addressing the challenges posed by traditional diagnostic methods. Utilizing the "Skin Cancer: Malignant vs. Benign" dataset, the hybrid model achieved an accuracy of 84%, with precision, recall, F1 score, and specificity recorded at 85%, 84%, 84%, and 83%, respectively. These results underscore the model's potential to surpass single-algorithm approaches in detecting skin cancer, making it a promising tool for early diagnosis and better-informed clinical decision-making. The findings highlight the broader impact of advanced machine learning techniques in healthcare, particularly in oncology, by demonstrating how the integration of multiple algorithms can provide more accurate, scalable, and reliable diagnostic solutions. This research opens avenues for further exploration of hybrid models as a means to advance AI-driven diagnostic technologies in medical fields, with potential applications across various types of cancer detection. The source code for this study is available through a public GitHub repository, fostering transparency and further innovation in the field.

Keywords- Hybrid Machine Learning; Skin Cancer; Convolutional Neural Networks; Support Vector Machines; and Gradient Boosting.

I. INTRODUCTION

Cancer continues to be a pressing global health concern, accounting for a significant share of mortality rates around the world [1][2]. Early detection and accurate diagnosis are vital components in enhancing patient outcomes [3][4], particularly in the case of skin cancer, where timely interventions can lead to markedly better survival chances [5]. While established diagnostic methods, such as visual inspection, biopsy, and histopathology have their merits, they are often susceptible to human error and subjectivity in interpretation [6]. These limitations highlight the urgent need for more reliable and automated diagnostic tools that can support healthcare professionals in making consistent and accurate diagnoses.

In the realm of healthcare, Machine Learning (ML) has emerged as a transformative force, providing innovative ways to enhance diagnostic accuracy by efficiently analyzing large and complex datasets. Various ML models, including Convolutional Neural Networks (CNNs) and Support Vector Machines (SVMs), have been successfully applied in cancer diagnosis [7][8]. However, each model comes with its own set of challenges. For instance, while CNNs excel at extracting meaningful features from image data, their effectiveness can diminish when training datasets are insufficient. On the other hand, SVMs are proficient at managing high-dimensional data but may struggle with scalability when confronted with larger datasets [9]. These nuances highlight the necessity of refining diagnostic methodologies to fully harness the potential of machine learning.

Hybrid machine learning models have gained prominence as a solution to the inherent limitations of individual techniques, leveraging the complementary strengths of multiple algorithms to enhance performance across a range of applications [10]. While hybridization itself is well-established, the contribution of this research lies in the sophisticated fusion of CNNs, SVMs, and Gradient Boosting (GB), each selected for their distinct advantages in the context of skin cancer diagnosis. CNNs are employed for their unparalleled ability to autonomously extract hierarchical features from complex image data, effectively capturing spatial patterns that are crucial for medical image analysis. SVMs, known for their robustness in high-dimensional spaces, are used to classify these features accurately, especially in cases with intricate decision boundaries. GB, recognized for its ensemble learning capability, is integrated to refine predictive accuracy, enhance model generalization, and mitigate the risks of overfitting.

The innovation in this approach extends beyond the selection of individual algorithms; it lies in their seamless integration through advanced fusion strategies, such as

ISBN: 978-1-68558-247-0

weighted averaging and voting. These techniques harmonize the outputs of the models, optimizing the decision-making process and enhancing the overall reliability of predictions. This methodological synergy not only strengthens the model's resilience to variations in the data but also ensures a robust and scalable solution for skin cancer diagnosis.

The effectiveness of the hybrid framework is demonstrated through its superior performance across key metrics, including accuracy, precision, recall, F1 score, and specificity, compared to models based on a single algorithm. This research provides a detailed exploration of the methodology, experimental design, and results, underscoring the potential of hybrid models to significantly advance diagnostic capabilities. By strategically combining CNNs, SVMs, and GB, this approach offers a novel solution that not only leverages the strengths of each algorithm but also mitigates their individual weaknesses, establishing a compelling contribution to the field of medical diagnostic systems.

The structure of the rest of this paper is organized as follows: Section II explores related works, emphasizing the various machine learning approaches employed in cancer diagnosis, hybrid algorithms used in cancer research, and the data types and sources referenced in prior studies. Section III outlines our research approach in detail, including descriptions of data collection and preprocessing, model development, and methodological flowcharts that illustrate the study's workflow. Section IV presents our findings, supported by a detailed analysis and contextual discussion to interpret their significance. Finally, Section V summarizes the key insights of the study, highlights its limitations, and offers directions for future research.

II. LITERATURE REVIEW

Prompt and accurate identification is crucial for effective treatment and improved patient results. Conventional methods for diagnosing cancer, including imaging, histology, and genetic testing, are restricted in their ability to accurately detect, identify, and understand the progression of the disease. Studies have shown that advancements in machine learning have significantly impacted various sectors, including healthcare and these developments have facilitated the design and implementation of automated diagnostic tools, which have demonstrated improved accuracy in specific applications, as evidenced by numerous studies and clinical trials [11].

A. *Machine Learning in Cancer Diagnosis*

There are four main types of machine learning: supervised, unsupervised, semi-supervised, and reinforcement learning. Supervised learning involves training models on labeled data for tasks like classification and regression, while unsupervised learning identifies patterns and structures within unlabeled data, often through clustering techniques. Semi-supervised learning combines a small amount of labeled data with a larger pool of unlabeled data to improve model performance, and reinforcement learning teaches agents to make sequential decisions by rewarding desired behaviours and penalizing undesired ones.

To support these types of learning, preprocessing techniques like feature selection and normalization are essential, as they improve model effectiveness by selecting the most relevant features and ensuring that data is appropriately scaled. These steps contribute to higher model accuracy and efficiency, particularly when dealing with diverse datasets across machine-learning approaches. In the field of cancer diagnosis, where data is often complex and high-dimensional, such preprocessing methods play a crucial role in optimizing model performance. Hybrid models, in particular, benefit from combining classification and clustering techniques to handle different aspects of the data [12]. For instance, cancer diagnosis hybrid models incorporate both classification and clustering methods and often integrate feature selection and dimensionality reduction techniques to further enhance diagnostic accuracy and efficiency.

Cluster-based classification involves using clustering algorithms to organize data before training a classifier, which helps to reduce variation within clusters and enhance overall performance. Combining predictions from various models like SVM, Random Forest, and Neural Networks in ensemble approaches enhances both accuracy and robustness. In addition, deep learning with feature engineering involves utilising deep learning models to extract features directly from raw data like images or genetic sequences and inputting them into standard classifiers for improved prediction [13].

As Artificial Intelligence (AI) and ML quickly grow in importance, they're becoming essential tools in healthcare, especially for diagnosing and treating diseases. Cancer diagnosis is a prime example where these technologies can make a real difference [14]. Growing evidence shows that detecting cancer at an early stage leads to better treatment options and significantly improved patient outcomes. Consequently, researchers have leveraged hybrid machine learning models, combining multiple algorithms to capitalize on their individual strengths and compensate for their limitations, thus enhancing diagnostic accuracy and robustness in cancer detection [15][16]. In cancer detection, for instance, ensemble techniques like bagging, boosting, or stacking are commonly used to blend predictions from multiple models, thereby enhancing accuracy and reliability. Deep learning models, such as CNNs, are often paired with traditional algorithms like SVMs, creating a robust approach especially valuable for complex data, such as medical imaging [17].

An additional strength of these hybrid models lies in their ability to integrate diverse data types which may include genomics, imaging, and clinical records into a unified diagnostic tool. By drawing on multiple sources, hybrid models can better capture the complex biological patterns associated with cancer. For example, logistic regression and elastic net techniques are sometimes applied to classify genetic variations with higher risks, while specialized tools like LungCLiP use ensemble classifiers to detect lung cancer in plasma samples [18]. Altogether, these integrative machine learning approaches are advancing diagnostic accuracy, supporting earlier detection, and enabling more personalized treatment strategies in cancer care [19].

B. *Hybrid Algorithms in Cancer Studies*

Hybrid algorithms in cancer research employ various techniques to enhance the precision and effectiveness of cancer diagnosis, treatment planning, and prognosis prediction. Selecting the right hybrid approach depends on multiple factors, including the type of cancer, the nature of available data, specific diagnostic objectives, and desired outcomes [20][21]. CNNs with SVMs for example takes advantage of both approaches' strengths: CNNs' excellent feature extraction capabilities and SVMs' ability to classify small numbers of data [22]. Saleh et al. [22] used a hybrid CNN-SVM method for classifying lung CT images into four categories: adenocarcinoma, large cell carcinoma, normal, and squamous cell carcinoma. The method was tested on the Chest CT-Scan images dataset, achieving a high classification accuracy rate of 97.91%, which outperformed other recent deep learning-based works.

The method used by Saleh et al. [22] also demonstrated promising performance in terms of sensitivity, specificity, precision, and AUC. The results suggest that the hybrid CNN-SVM method has the potential to assist in the early detection of lung cancer, and future work can focus on testing the method with different datasets and image types [22]. Combining CNN and Recurrent Neural Networks (RNNSs) for cancer diagnosis as demonstrated in [23] takes advantage of the capabilities of each architecture and in [8] where CNNs extract spatial features, whereas RNNs recognize temporal or sequential patterns. This hybrid technique is particularly useful for evaluating complicated medical data, such as time-series medical pictures or sequential patient records.

These studies highlight some of the advancements in cancer diagnosis through hybrid machine learning models. By combining the strengths of different algorithms, such as CNNs with SVMs or RNNs, hybrid approaches have demonstrated improved performance in various cancer detection tasks, achieving higher accuracy and reliability compared to traditional single-algorithm methods. Table I shows reviews of related works that use Hybrid Machine Learning for Cancer Diagnosis.

TABLE I. REVIEW OF OTHER RELATED STUDIES

Article Ref.	**Data Source**	**Records**	**Train/Test Split**	**Algorithm Type**	**Model Accuracy**
[22]	Chest CT-Scan Images	5103	80:20	CNN, SVM	97.91%
[29]	Herlev public	917	80:20	CNN, SVM	99.30%
[31]	Breast Cancer Network Wisconsin	-	70-30	K-means, SVM	97.34%
[30]	Mammographic Image Analysis Society	-	70-30%	CNN, GRU	95.50%
[32]	PCAM Kaggle	277524	80:20	CNN, GRU	86.21%
[23]	-Lung Image Database Consortium and Image Database Resource Initiative (LIDC-IDRI)	888	80:20	CNN, RNN	95.00%

C. *Datasets and Data Sources*

Datasets are the foundation of machine learning, essential for training, validating, and testing models to ensure robustness, generalizability, and the ability to tackle real-world challenges effectively. A well-curated, diverse, and high-quality dataset is crucial to the success of any machine learning effort, as it enables models to perform accurately while minimizing biases. For example, Yogendra Singh Solanki et al. [24] developed an ML-based classifier system for breast cancer prognosis using a dataset from the University of California, Irvine (UCI) repository to distinguish between malignant and benign breast cancer cells. In building such models, data imbalance often poses a significant challenge, as it can skew predictions toward the more prevalent class. Yogendra Singh Solanki et al. addressed this by using techniques like re-sampling and the Synthetic Minority Over-sampling Technique (SMOTE), a method for handling class imbalance by generating new realistic samples for the minority class, which helps models learn more evenly and reduces bias toward the majority class.

In their study, Wang et al. [25] obtained cancer data from three distinct data sources to analyze cancer incidences, which include, The Cancer Genome Atlas (TCGA), Surveillance, Epidemiology, and End Results (SEER) 18, and North American Association of Central Cancer Registries. The TCGA shared information on individuals with 33 different forms of cancer, using specific TCGA case IDs to prevent any repeat cases among the different types of cancer. SEER data includes individuals who were diagnosed with primary cancer between 2010 and 2013, characterized by the third edition of the International Classification of Diseases for Oncology (ICD-O-3) using primary site and histology/behavior criteria. The NAACCR database included cancer records from every state in the US as well as the District of Columbia, encompassing nearly the entire population of the country from 2009 to 2013.

In order to analyze the distribution of races, only cancer cases in the US with race information were used, taking into account the SEER program's overrepresentation of minority populations in the US. These expansive data collections cover 33 different types of cancer and can be identified by distinct case IDs and ICD-O-3 categorization, offering extensive population representation and valuable insights into cancer case characteristics in the US [25].

III. METHODOLOGY

This paper focuses on developing a robust hybrid machine learning model for cancer diagnosis, encompassing several key phases: data collection and preprocessing, exploratory data analysis, feature engineering, model selection, and implementation following our methodology framework as in Figure 1. Utilizing the "Skin Cancer: Malignant vs. Benign" image dataset, exploratory data analysis provided insights into the dataset composition, guiding feature engineering, which tailored the data for use in CNNs, SVMs, and GB models. Model selection and implementation involved carefully combining these techniques using Python and Scikit-Learn, optimizing the model's performance through hyperparameter tuning and an 80/20 train-test split. The results demonstrated the model's effectiveness, offering promising advancements in accurate cancer diagnosis.

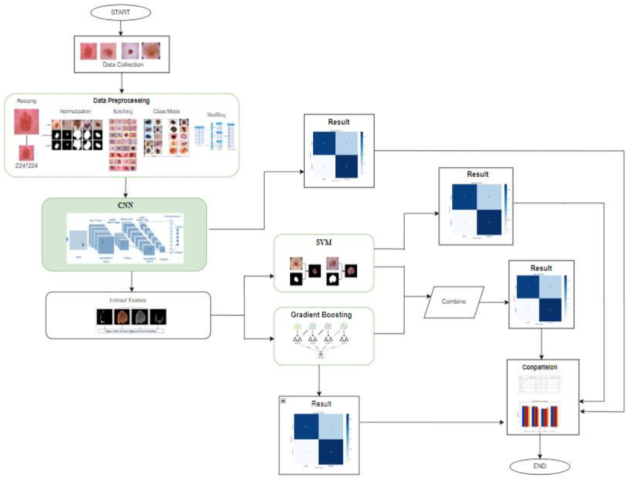


Figure 1. Research methodology flowchart.

A. *Data Collection and Preprocessing*

The "Skin Cancer: Malignant vs Benign" dataset [26], used in this study, consists of 3,600 images of skin moles, evenly divided between benign (1,800) and malignant (1,800) cases. Each image, sourced from the ISIC Archive, a well-regarded repository for dermatological images is provided at a consistent resolution of 224x244 pixels, ensuring uniform quality for analysis. This balanced dataset with samples shown in Figure 2 below serves as an essential resource for developing and validating machine learning models to improve the accuracy and reliability of skin cancer diagnosis.

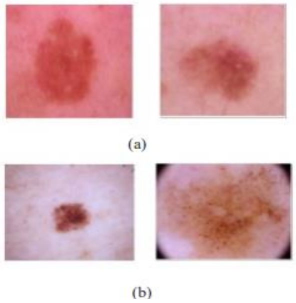


Figure 2. Skin cancer Images- (a) Benign cancer, (b) Malignant cancer

With a training set of 2,637 images (1,440 benign and 1,197 malignant) and a test set of 963 images (360 benign and 300 malignant), split into an 80/20 ratio for training and testing, this dataset provides a robust foundation for training and evaluating hybrid machine learning models for cancer diagnosis.

The boxplot in Figure 3 and Figure 4 visualizes the distribution of the "mean intensity" feature for benign and malignant tumor images based on the training and test data respectively, highlighting the differences in typical values (median, quartiles) and variability (spread) between the two classes. Outliers, represented as individual circles outside the whiskers, indicate data points that fall significantly outside the general distribution. This suggests that mean intensity could be a valuable feature for classifying tumors, as benign tumors generally show lower mean intensity compared to malignant ones. Additionally, this feature could be used directly in machine learning models or serve as a basis for creating more complex features. Handling outliers may be necessary to avoid skewing model training, and understanding the distribution of mean intensity can inform decisions on data preprocessing and feature engineering strategies.

The images were resized to 224x224 pixels to ensure uniformity and compatibility with CNN. The pixel values are normalized to a range between 0 and 1 to improve model convergence during training. The dataset is also shuffled to prevent overfitting, ensuring that the model learns from a balanced and randomized distribution of benign and malignant cases.

CNNs are employed for automatic feature extraction from the image data. In parallel, the Gabor filters and Gray-Level Co-occurrence Matrix (GLCM) are applied for texture-based feature extraction. The extracted features from both methods are then combined through feature fusion to enhance classification capabilities. Principal Component Analysis (PCA) was then used to reduce the dimensionality of the combined feature set. This step was to ensure that the most important features are retained while reducing computational complexity, leading to a more efficient and scalable model.

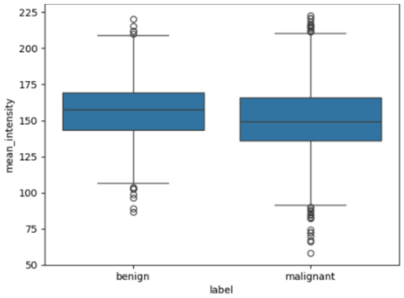


Figure 3. Boxplot of mean intensity by label (Training Data)

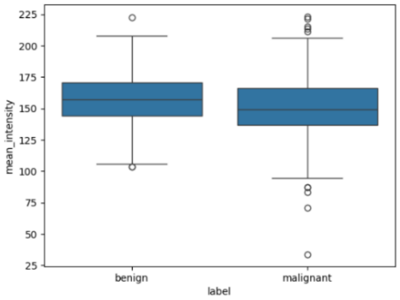


Figure 4. Boxplot of mean intensity by label (Test Data)

B. *Model Development*

This study develops a robust hybrid machine learning model for skin cancer diagnosis, combining the strengths of CNNs, SVMs, and GB. The model's core design is based on leveraging the complementary capabilities of these

ISBN: 978-1-68558-247-0

algorithms: CNNs for automatic feature extraction from skin cancer images, SVMs for efficient classification in high-dimensional feature spaces, and GB for ensemble learning that refines predictive performance by aggregating weak learners into a more robust model. Specifically, CNNs excel at identifying intricate spatial patterns within the images, while SVMs handle complex decision boundaries in the high-dimensional feature space, and GB enhances generalization by reducing overfitting.

The integration of these algorithms is facilitated by advanced fusion strategies, such as weighted averaging and voting, which harmonize the individual outputs to optimize decision-making. This methodological synergy ensures that the model's predictive reliability is enhanced, particularly in clinical settings where diagnostic accuracy is paramount. The integration process is sequential, with CNNs first extracting relevant features, followed by SVM classification and GB aggregation, to ensure a comprehensive approach to skin cancer diagnosis.

The model was implemented using Python 3.12.7 and the following libraries: Scikit-learn 1.4.2 for SVM and GB, Keras/TensorFlow for building and training the CNN, and Pandas, NumPy, Matplotlib, and Seaborn for data manipulation, feature engineering, and visualization. The "Skin Cancer: Malignant vs. Benign" dataset, sourced from the ISIC Archive, comprises 3,600 images, evenly distributed between benign and malignant cases. Images are resized to 224x224 pixels and normalized to improve model convergence during training. Data preprocessing includes shuffling to prevent overfitting and feature engineering through Gabor filters and Gray-Level Co-occurrence Matrix (GLCM), with dimensionality reduction applied using PCA to maintain computational efficiency while retaining essential features.

Hyperparameter tuning was performed on each model component: CNN layers, SVM parameters, and GB hyperparameters, using grid search and cross-validation to optimize performance. Evaluation metrics, including accuracy, precision, recall, F1 score, and specificity, were calculated and compared to single-algorithm models to demonstrate the efficacy of the hybrid approach. The confusion matrix was also generated to visualize the model's classification performance in terms of true positives, false positives, true negatives, and false negatives.

The dataset was partitioned into an 80% training set and a 20% testing set, ensuring that the model underwent thorough training while maintaining an unbiased and reliable evaluation process. This division facilitates both robust model learning and accurate performance assessment. The hybrid model's design effectively addresses several critical challenges commonly encountered in medical diagnostics, such as data sparsity, class imbalance, and computational scalability. By leveraging the strengths of multiple algorithms, the model not only mitigates the limitations of individual approaches but also achieves superior diagnostic performance, outperforming single-algorithm models in terms of accuracy and generalization. The source code for the implementation is publicly accessible via a GitHub repository [27], fostering transparency and providing an avenue for further research and development within the domain of skin cancer detection. This open-access model serves as a valuable resource for advancing the field and promoting collaborative exploration of hybrid machine learning techniques for medical image analysis.

IV. RESULTS AND DISCUSSION

In this study, a thorough assessment and comparison of various machine learning models for cancer detection were carried out using SciKit-Learn. Table II shows the performance metrics Accuracy, Precision, Recall, F1-score, and Specificity for different models including the proposed hybrid model, CNN, SVM, and GB.

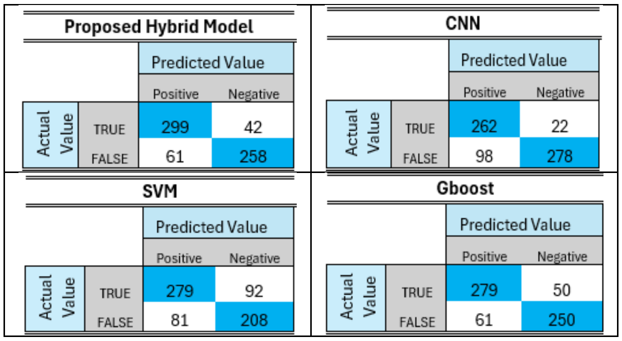
TABLE II. ML MODELS PERFORMANCES DATA

Metric	Hybrid Model	CNN	SVM	GB
Accuracy	84	82	74	83
Precision	85	84	74	82
Recall	84	82	74	85
F1	84	82	74	83
Specificity	83	73	78	80

Table III below shows confusion matrices for the Hybrid, CNN, GB, and SVM models illustrating the breakdown of True Positives (TP), False Positives (FP), True Negatives (TN), and False Negatives (FN) values essential for evaluating a model's classification performance by breaking down its accurate and inaccurate predictions, offering a comprehensive view of its classification abilities [28] while Figure 5 offers a comparison of these performance measures among the various models.

The findings of this study underscore the significance of the hybrid machine learning model in the field of skin cancer diagnostics. The model demonstrated high accuracy and robust performance metrics, suggesting its potential utility as a reliable tool for the early detection of skin cancer.

TABLE III. CONFUSION MATRIX OF PROPOSED HYBRID MODEL, CNN, GB, AND SVM MODELS.



Proposed Hybrid Model

		Predicted Value	
		Positive	Negative
Actual Value	TRUE	299	42
	FALSE	61	258

CNN

		Predicted Value	
		Positive	Negative
Actual Value	TRUE	262	22
	FALSE	98	278

SVM

		Predicted Value	
		Positive	Negative
Actual Value	TRUE	279	92
	FALSE	81	208

Gboost

		Predicted Value	
		Positive	Negative
Actual Value	TRUE	279	50
	FALSE	61	250

By properly evaluating skin cancer photos and utilizing the characteristics of many algorithms, the model has proved its capacity to reliably discern between benign and malignant skin moles. This is a step forward in the development of more precise diagnostic tools, perhaps leading to earlier identification and improved treatment outcomes for patients.

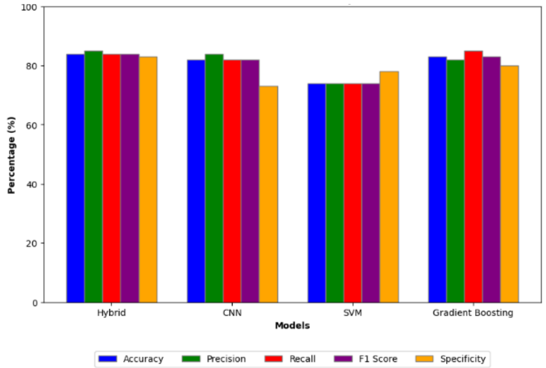


Figure 5. Visualization of the model's performances

The findings of this study underscore the significance of the hybrid machine learning model in the field of skin cancer diagnostics. The model demonstrated high accuracy and robust performance metrics, suggesting its potential utility as a reliable tool for the early detection of skin cancer. By properly evaluating skin cancer photos and utilizing the characteristics of many algorithms, the model has proved its capacity to reliably discern between benign and malignant skin moles. This is a step forward in the development of more precise diagnostic tools, perhaps leading to earlier identification and improved treatment outcomes for patients.

The model's strong performance in skin cancer diagnosis shows that it might be useful in clinical settings, allowing

Medical practitioners make better informed decisions and eliminate diagnostic mistakes. Overall, the model's contribution to improving skin cancer detection is important for successful treatment and increasing patient survival rates.

V. CONCLUSION

This research focused on evaluating the effectiveness of a hybrid machine learning model for skin cancer diagnosis compared to traditional models like CNN, SVM, and GB. The hybrid model demonstrated superior performance with an accuracy of 84%, precision of 85%, recall of 84%, F1 score of 84%, and specificity of 83%. These metrics indicate that the hybrid model not only performed better than the SVM model, which had the lowest accuracy at 74%, but also outperformed the CNN and Gradient Boosting models in most aspects, particularly in specificity. The CNN model, while achieving high precision (84%) and accuracy (82%), lagged in specificity (73%), indicating a higher rate of false positives. In contrast, the Gradient Boosting model, with metrics closely matching the hybrid model, also showed strong performance but was slightly less effective overall.

This summary highlights the hybrid model's potential as a more reliable and accurate tool for skin cancer diagnosis. This paper is primarily focused on comparing single-algorithm models to hybrid Machine Learning models, however for future research directions, the proposed hybrid models can be compared with other potential hybrid models.

REFERENCES

[1] H. Sung et al., "Global cancer statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries," CA Cancer J Clin, vol. 71, pp. 209–249, May 2021, doi: 10.3322/caac.21660.

[2] M. Dildar et al., "Skin cancer detection: A review using deep learning techniques," Int. J. Environ. Res. Public Health, vol. 18, pp. 5479, May 2021, doi: 10.3390/ijerph18105479.

[3] Y. Xie et al., "Early lung cancer diagnostic biomarker discovery by machine learning methods," Translational oncology., vol. 14, pp100907, Jan. 2021, doi: 10.1016/j.tranon.2020.100907.

[4] X. Guan et al., "Construction of the XGBoost model for early lung cancer prediction based on metabolic indices," BMC medical informatics and decision making, vol. 23, pp. 107, Jun. 2023, doi: 10.1186/s12911-023-02171-x.

[5] A. Imran, A. Nasir, M. Bilal, G. Sun, A. Alzahrani, and A. Almuhaimeed, "Skin cancer detection using combined decision of deep learners," IEEE Access, vol. 10, pp. 118198–118212, Nov. 2022, doi: 10.1109/ACCESS.2022.3220329.

[6] A. Echle, N. T. Rindtorff, T. J. Brinker, T. Luedde, A. T. Pearson, and J. N. Kather, "Deep learning in cancer pathology: a new generation of clinical biomarkers," British journal of cancer, vol. 12, pp. 686-696, Feb. 2021, doi: 10.1038/s41416-020-01122-x.

[7] G. Alfian et al., "Predicting breast cancer from risk factors using SVM and Extra-Trees-based Feature Selection Method," Computers, vol. 11, pp. 136, Sep. 2022, doi: 10.3390/computers11090136.

[8] Y. Lu et al., "A hybrid CNN-RNN approach for survival analysis in a Lung Cancer Screening study," Heliyon, vol. 9, pp. e18695, Aug. 2023, doi: 10.1016/j.heliyon. 2023.e18695.

[9] S. Alghunaim and H. H. Al-Baity, "On the scalability of machine-learning algorithms for breast cancer prediction in big data context," *IEEE Access*, vol. 7, pp. 91535–91546, Jun. 2019, doi: 10.1109/ACCESS.2019.2927080.

[10] A. Sahu, P. K. Das, and S. Meher, "High accuracy hybrid CNN classifiers for breast cancer detection using mammogram and ultrasound datasets," Biomed Signal Process Control, vol. 80, pp. 104292, Feb. 2023, doi: 10.1016/j.bspc.2022.104292.

[11] M. Javaid, A. Haleem, R. Pratap Singh, R. Suman, and S. Rab, "Significance of machine learning in healthcare: Features, pillars and applications," International Journal of Intelligent Networks, vol. 3, pp. 58–73, Jan. 2022, doi: 10.1016/j.ijin.2022.05.002.

[12] S. Bouhsissin, N. Sael, F. Benabbou, and A. Soultana, "Enhancing machine learning algorithm performance through feature selection for driver behavior classification," Indonesian Journal of Electrical Engineering and Computer Science, vol. 35, pp. 354, Jul. 2024, doi: 10.11591/ijeecs.v35.i1.pp354-365.

[13] X. Guo, M. Wei, H. Jiang, Z. Feng, and Y. Sun, "Data feature extraction methods based on deep learning," 2024 IEEE 4th International Conference on Electronic Technology, Communication and Information (ICETCI), IEEE, May 2024, pp. 1485–1489. doi: 10.1109/ICETCI61221.2024.10594224.

[14] M. J. Iqbal et al., "Clinical applications of artificial intelligence and machine learning in cancer diagnosis: looking into the future," Cancer Cell Int., vol. 21, pp. 270, May 2021, doi: 10.1186/s12935-021-01981-1.

[15] R. Chandragiri, D. B. Reddy, K. S. P. Badarla, A. H. Syed, K. Modepalli and K. Sadam, "A systematic review on machine learning techniques used for early detection of skin cancer," 2024 IEEE 13th International Conference on Communication Systems and Network Technologies (CSNT), IEEE, Jun. 2024, pp. 862–867, doi: 10.1109/CSNT.2024.140.

[16] M. Banda, E. K. Ngassam, and E. Mnkandla, "Enhancing classification and prediction through the application of hybrid machine learning models," 2024 IST-Africa Conference (IST-Africa), IEEE, May 2024, pp. 1–12. doi: 10.23919/IST-Africa63983.2024.10569590.

[17] S. Kusuma, S. G. Krishnan, K. Samreen, M. V. Ramana, and G. S. Prasad, "A hybrid deep learning approach for early detection and classification of lung cancer using the pelican optimization algorithm," 2024 International Conference on Distributed Computing and Optimization Techniques (ICDCOT), IEEE, Mar. 2024, pp. 1–6. doi: 10.1109/ICDCOT61034.2024.10515355.

[18] M. A. Zeller, Z. W. Arendsee, G. J. D. Smith, and T. K. Anderson, "classLog: Logistic regression for the classification of genetic sequences," Frontiers in Virology, vol. 3, pp. 1215012, Dec. 2023, doi: 10.3389/fviro.2023.1215012.

[19] K. Swanson, E. Wu, A. Zhang, A. A. Alizadeh, and J. Zou, "From patterns to patients: Advances in clinical machine learning for cancer diagnosis, prognosis, and treatment," Cell, vol. 8, pp. 186, Apr. 2023, doi: 10.1016/j.cell.2023.01.035.

[20] S. Molaei, S. Cirillo, and G. Solimando, "Cancer detection using a new hybrid method based on pattern recognition in micrornas combining particle swarm optimization algorithm and artificial neural network," Big Data and Cognitive Computing, vol. 8, pp. 33, Mar. 2024, doi: 10.3390/bdcc8030033.

[21] O. Khouadja, M. Jemai, and M. S. Naceur, "Improving cancer diagnosis with hybrid neural networks: A comprehensive narrative review," 2024 IEEE International Conference on Advanced Systems and Emergent Technologies (IC_ASET), IEEE, Apr. 2024, pp. 1–6. doi: 10.1109/IC_ASET61847.2024.10596187.

[22] A. Y. Saleh, C. K. Chin, V. Penshie, and H. R. H. Al-Absi, "Lung cancer medical images classification using hybrid cnn-svm," International Journal of Advances in Intelligent Informatics, vol. 7, pp. 151–162, Jul. 2021, doi: 10.26555/ijain.v7i2.317.

[23] S. Wankhade and S. Vigneshwari, "A novel hybrid deep learning method for early detection of lung cancer using neural networks," Healthcare Analytics, vol. 3, pp. 100195, Nov. 2023, doi: 10.1016/j.health.2023.100195.

[24] Y. S. Solanki et al., "A hybrid supervised machine learning classifier system for breast cancer prognosis using feature selection and data imbalance handling approaches," Electronics (Basel), vol. 10, pp. 699, Mar. 2021, doi: 10.3390/electronics10060699.

[25] X. Wang et al., "Characteristics of the cancer genome atlas cases relative to U.S. general population cancer cases," Br J Cancer, vol. 119, pp. 885–892, Oct. 2018, doi: 10.1038/s41416-018-0140-8.

[26] C. Fanconi, "Skin cancer: Malignant vs. Benign," ISIC archive Kaggle dataset. Accessed: Jan. 30, 2025. [Online]. Available: https://www.kaggle.com/datasets/fanconic/skin-cancer-malignant-vs-benign.

[27] Fazila Patel, "HybridML-SkinCancerDiagnosis, GitHub," 2024. Accessed: Jan. 31, 2025. [Online]. Available: https://github.com/patelfazila/HybridML-SkinCancerDiagnosis.

[28] L.-E. Pomme, R. Bourqui, R. Giot, and D. Auber, "Relative confusion matrix: Efficient comparison of decision models," 2022 26th International Conference Information Visualisation (IV), IEEE, Jul. 2022, pp. 98–103. doi: 10.1109/IV56949.2022.00025.

[29] A. Dongyao Jia, B. Zhengyi Li, and C. Chuanwang Zhang, "Detection of cervical cancer cells based on strong feature CNN-SVM network," Neurocomputing, vol. 411, pp. 112–127, Oct. 2020, doi: 10.1016/j.neucom.2020.06.006.

[30] P. E. Jebarani, N. Umadevi, H. Dang, and M. Pomplun, "A novel hybrid K-Means and GMM machine learning model for breast cancer detection," IEEE Access, vol. 9, pp. 146153–146162, Oct. 2021, doi: 10.1109/ACCESS.2021.3123425.

[31] B. Zheng, S. W. Yoon, and S. S. Lam, "Breast cancer diagnosis based on feature extraction using a hybrid of K-means and support vector machine algorithms," Expert Syst Appl, vol. 41, pp. 1476–1482, 2014, doi: 10.1016/j.eswa.2013.08.044.

[32] X. Wang et al., "Intelligent hybrid deep learning model for breast cancer detection," Electronics (Switzerland), vol. 11, Sep. 2022, doi: 10.3390/electronics11172767.

ISBN: 978-1-68558-247-0