IMPROVING HEART DISEASE DETECTION AND PREDICTION THROUGH DEEP AND MACHINE LEARNING ALGORITHMS IN COMPUTED TOMOGRAPHY (CT) IMAGERY

Zacheous kasonda [1*], Dr.R. Priscilla[2], Dr.Glorindal Selvam[3]

[1*] Technical Manager, DMI-St Eugene University, Zambia
zkasonda@gmail.com
[2] Professor, St. Joseph's Institute of Technology, Chennai
prisci.christa@gmail.com
[3] Professor, DMI-St Eugene University, Zambia
glorygj@yahoo.com

ABSTRACT
The aim of this study is to investigate the effectiveness of deep and machine learning algorithms for detecting and predicting heart disease using CT imagery in Malawi, and to compare their performance with traditional methods. The study involves data collection and pre-processing, model development and testing, and comparison of the performance of various algorithms. Through this study, to gain a better understanding of the strengths and limitations of deep and machine learning algorithms in this critical medical application.

Keywords: Deep learning algorithm, CT imagery, Heart disease, Medical imaging, Diagnosis, Machine learning algorithms

INTRODUCTION

Heart disease is a global health concern affecting millions of people. Early detection and diagnosis are critical for effective treatment and prevention of complications. Deep learning and machine learning algorithms present a promising approach for heart disease detection. The proposed system for heart disease detection utilizes CT imagery in resource-limited settings, where access to specialized diagnostic tools and expertise is limited. The system allows doctors and healthcare professionals to upload CT scans, which are then prepared and analyzed using deep and machine learning algorithms to identify potential signs of coronary artery disease. The resulting report provides relevant findings to support informed decision- making about patient care. This system has the potential to be an effective diagnostic tool in the fight against heart disease.

1. To evaluate and compare the performance of deep and machine learning algorithms in thedetection and prediction of heart disease using CT imagery.
2. To identify the algorithm with the highest prediction and detection rate for diagnosingheart disease using CT imagery and present an outline of its performance.

In recent years, deep learning algorithms have been gaining attention in the medical field for their ability to accurately detect complex patterns and structures in medical images, such as CT scans [2]. Studies have shown that deep learning algorithms, particularly Convolutional Neural Networks (CNNs), can accurately identify abnormalities and lesions in CT images, such as lung nodules [4] and breast cancer [5]. CNNs provide more detailed feature extraction from images and automatically classify features for better identification, resulting in high levels of accuracy for detecting and predicting heart disease using CT imagery [6].

In addition to deep learning, machine learning techniques have also been used for heart disease prediction [8][26]. Machine learning allows algorithms to learn from data and improve their performance over time, making it well-suited for predicting heart disease by identifying patterns and trends in large and complex datasets. A study using a decision tree-based algorithm showed better performance in predicting heart disease than logistic regression and artificial neural networks [3][21]. Further research has since been conducted to develop more sophisticated machine learning algorithms, such as the use of support vector machines (SVMs) and KNN [10]. These methods hold great promise for improving the detection and prevention of heart disease, but their performance can vary depending on factors such as dataset quality, specific techniques used, and algorithm parameters[22]. Therefore, continued research is needed to fully understand and develop more effective algorithms for heart disease prediction [8].

THEORETICAL FRAMEWORK

The artificial intelligence (AI) theoretical framework for heart disease prediction and detection has been widely studied in the literature. These frameworks typically involve the use of machine and deep learning techniques, such as neural networks [11], to develop predictive models for heart disease[20]. This is the key theoretical approach. One popular approach under the AI framework is the use of convolutional neural networks (CNNs) [11] for image classification and feature extraction from medical images, such as CT or MRI scans[23]. Overall, AI-based theoretical frameworks have demonstrated promising results in the literature for heart disease prediction and detection [12] and have the potential to improve the diagnosis and treatment of heart disease[18].

Model building phase where one or more deep and machine learning algorithms are trained on pre-processed data and the best-performing algorithm is selected using cross-validation[24][25]. Three algorithms for the model application selected are Support Vector Machine (SVM) [13],Random Forests, and Convolutional Neural Networks (CNNs) [14,15,16]. CNN is a popular algorithm for deep learning [19][27].

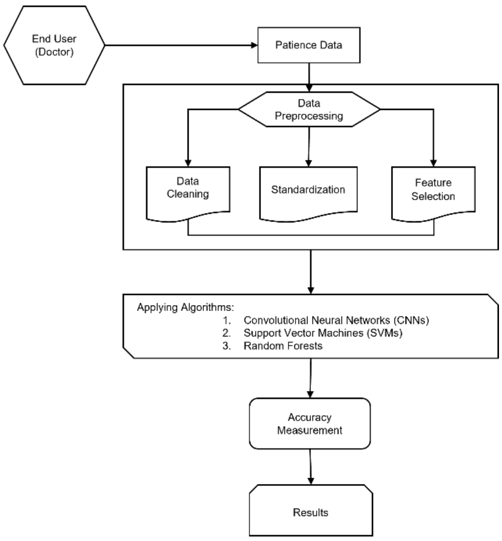


Fig.1: Proposed System Architecture

In the literature, there has been considerable research into the use of artificial intelligence (AI) frameworks to predict and detect heart disease. These frameworks often rely on machine and deep learning techniques, such as neural networks [11], to develop predictive models for heart disease. A popular AI approach is to use convolutional neural networks (CNNs) [11] for feature extraction and image classification in medical images, such as CT or MRI scans. The effectiveness of these AI-based frameworks has shown promise in the literature for improving the diagnosis and treatment of heart disease [12].

During the model building phase, one or more deep and machine learning algorithms are trained on preprocessed data, and the best-performing algorithm is selected using cross- validation. Three algorithms selected for the application of the model are Support Vector Machine (SVM) [13], Random Forests, and Convolutional Neural Networks (CNNs) [14,15,16]. Among these algorithms, CNNs are a popular choice for deep learning.

METHODOLOGY DATA COLLECTION

The research method of choice was the quantitative approach. The CT imagery heart dataset was obtained from the Kaggle.com repository. The data was utilised to test the model for heart disease detection and prediction using the algorithms. The key dataset characteristics were the two groups of images, which were images with and without heart disease. The images were provided in two folders train and test.

DATA STANDARDIZATION

Data was standardised using image processing techniques to resize and normalize the images, ensuring they were consistent in size and resolution (refer to Figure 2). Additionally, the system checked for any errors or inconsistencies in the images to ensure high-quality data suitable for analysis.

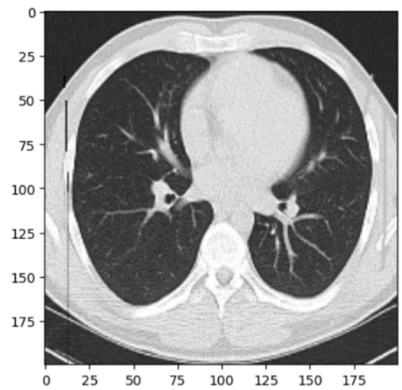


Fig 2: Visualising matplot data

To train and test the Random Forest, SVM, and CNN algorithms on the CT imagery heart dataset, the following steps were taken using Python:

- **Data preprocessing:** The dataset was loaded and preprocessed using image processing techniques to standardize the images' size and resolution.
- **Feature extraction:** The key features were extracted from the images to improve the algorithms' performance.
- **Model training:** The preprocessed data was used to train the machine learning algorithms, which included the Random Forest and SVM algorithms.
- **Model selection:** The algorithms were evaluated using cross-validation, and the best-performing algorithm was selected.
- **Hyperparameter tuning:** The selected algorithm's hyperparameters were optimizedto obtain the best performance.
- **Model testing:** The selected algorithm was used to make predictions on the testdataset,

and its performance was evaluated.

- **Evaluation metrics:** The algorithms' performance was evaluated using accuracy, precision, recall, and F1-score metrics.

1. Dataset and split it into training and testing sets.
2. Train a Random Forest model on the training set.
3. Train an SVM model on the training set.
4. Train a CNN model on the training set.
5. Evaluate the performance of both models using accuracy, sensitivity, precision, specificity, confusion matrix, and ROC curves.

By following these steps, high levels of accuracy were achieved for the detection and prediction of heart disease using the CT imagery dataset.

Random Forest Model

The Random Forest algorithm is a machine learning technique that employs multiple decision trees to construct a more reliable and precise model. It operates by selecting a random subset of features and constructing several decision trees based on these features. The final prediction is made by combining the predictions of all these trees. Compared to a single decision tree, the Random Forest algorithm is less prone to overfitting, which is a significant advantage.

For this study, a Random Forest model with a maximum depth of 5 will be trained, and the testing set will be used to predict the target variable. The performance of the model will be assessed using accuracy, sensitivity, precision, specificity, confusion matrix, and ROC curves, which can be seen in Fig.3

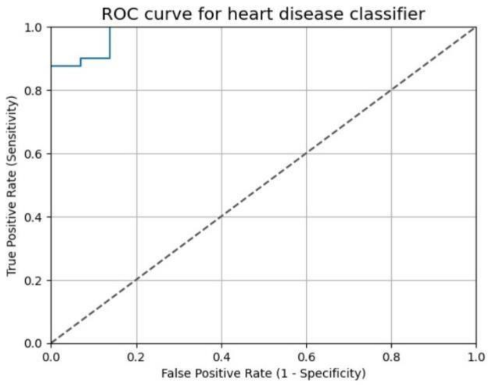


Fig.3: ROC curve for heart disease classifier

Support Vector Machine Model

The SVM algorithm is a discriminative model that separates data points using a hyperplane. The algorithm works by finding the hyperplane that maximizes the margin between the two classes. The SVM algorithm has the advantage of being effective in high-dimensional spaces and being less prone to overfitting than other models. In this analysis, we will train an SVM model using a linear kernel. We will then use the model to predict the target variable of the testing set. We evaluated the performance of the model using accuracy, sensitivity, precision, specificity, confusion matrix, and ROC curves.

CNN (Convolutional Neural Network)

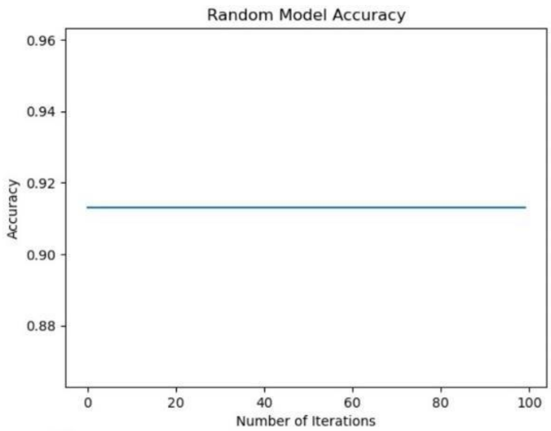
The CNN algorithm is a powerful deep learning method used for image classification and feature extraction from medical images such as CT scans. It works by processing the input image through a series of convolutional layers, pooling layers, and fully connected layers. This allows the network to learn complex features and patterns within the image data,improving accuracy and performance. In this analysis, we will train a CNN model with multiple convolutional layers and pooling layers.

BioGecko

Vol 12 Issue 01 2023
ISSN NO: 2230-5807

RESULTS

The Random Forest algorithm obtained an accuracy rate of 91.1% (see Figure 4), along with a sensitivity of 90%, precision of 94.7%, and specificity of 93.1%. Moreover, the Random Forest model generated Figure 5, which visualizes the prediction results of the tested images.



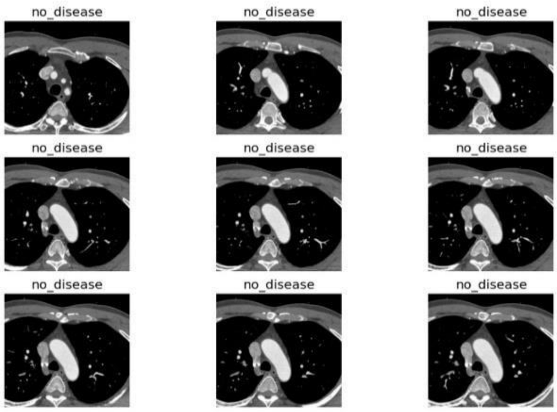
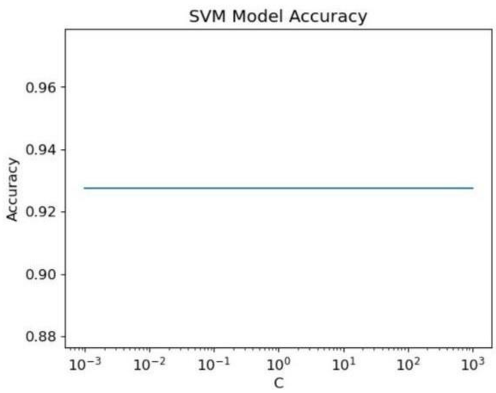


Fig 4: Random Forest model accuracy

Fig 5: Random Forest algorithm results

The accuracy achieved by the SVM model was 92.8% (as shown in Figure 6), with a sensitivity of 92.5%, a precision of 94.9%, and a specificity of 93.1%. The prediction results were also presented in the form of an image (Figure 7) generated by the SVM model.



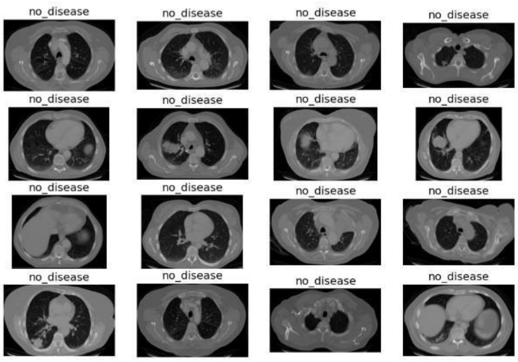


Fig 6: SVM model accuracy

Fig 7: SVM algorithm results

The CNN model yielded an accuracy of 98.6% (as shown in Figure 8), with a sensitivity of 95%, precision of 100%, and specificity of 100%. The model's prediction results are presented in Figure 9.

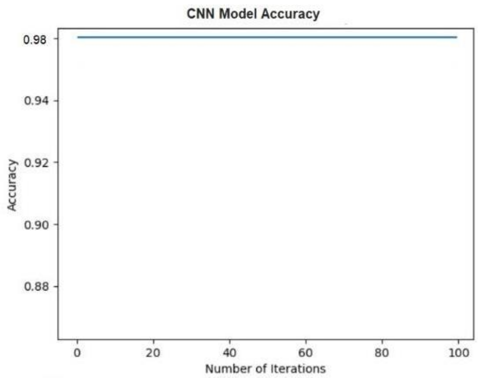


Fig 1: CNN model accuracy

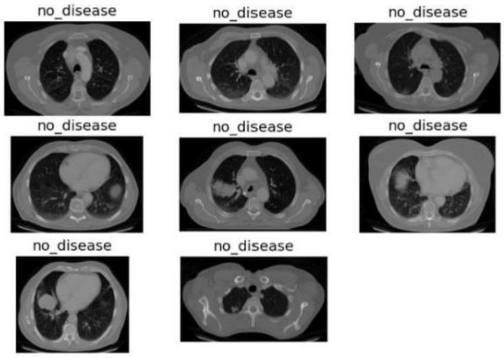


Fig 8: SVM algorithm results

Compared to the machine learning algorithms, the Convolutional Neural Network (CNN) achieved significantly higher accuracy in detecting and predicting heart disease, with an accuracy of 98.6% (refer to Figure 9). In contrast, the Random Forest model achieved an accuracy of 91.1% and the SVM achieved an accuracy of 92.8%.

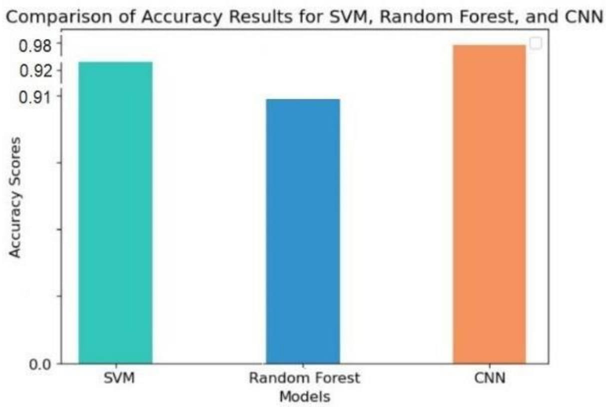


Fig 2: Accuracy Results

CONCLUSION

In conclusion, both Random Forest and SVM models were effective in detecting and predicting heart disease, but the CNN model achieved higher performance metrics. While the differences in accuracy are minor, the choice of algorithm may depend on the specific application. Therefore, a combination of all three algorithms could potentially increase accuracy in heart disease detection and prediction. The study also found that increasing the number of iterations improved the accuracy of all three models.

REFERENCES

1. Riyaz, Lubna, et al. "Heart disease prediction using machine learning techniques: a quantitative review." International Conference on Innovative Computing and Communications: Proceedings ofICICC 2021, Volume 3. Springer Singapore, 2022.
2. L. Mary Gladence, T. Ravi and M. Karthi, "An enhanced method for detecting congestive heart failure- Automatic Classifier," 2014 IEEE International Conference on Advanced Communications, Control and Computing Technologies, Ramanathapuram, India, 2014, pp. 586-590, doi: 10.1109/ICACCCT.2014.7019154.
3. Karthi, M., L. Mary Gladence, and J. Joshua. "A Comparative Study of Detecting the Congenital Heart Disease Using Tree Algorithms." RESEARCH JOURNAL OF PHARMACEUTICAL BIOLOGICAL AND CHEMICAL SCIENCES 7.4 (2016): 78-84.

4. Kumar, M. Kiran, and Anthoniraj Amalanathan. "Automated lung nodule detection in CT images by optimized CNN: impact of Improved whale optimization algorithm." Computer Assisted Methods in Engineering and Science 29.1–2 (2022): 7-31.
5. Celard, P., et al. "A survey on deep learning applied to medical images: from simple artificial neural networks to generative models." Neural Computing and Applications 35.3 (2023): 2291-2323.
6. Jafari, Mahboobeh, et al. "Automated Diagnosis of Cardiovascular Diseases from Cardiac Magnetic Resonance Imaging Using Deep Learning Models: A Review." arXiv preprint arXiv:2210.14909 (2022).
7. Sharif, Hanan, Faisal Rehman, and Amina Rida. "Deep Learning: Convolutional Neural Networks for Medical Image Analysis-A Quick Review." 2022 2nd International Conference on Digital Futures and Transformative Technologies (ICoDT2). IEEE, 2022.
8. Gladence, L. Mary, M. Karthi, and V. Maria Anu. "A statistical comparison of logistic regression and different Bayes classification methods for machine learning." ARPN Journal of Engineering and Applied Sciences 10.14 (2015): 5947-5953.
9. Gladence, L. Mary, M. Karthi, and T. Ravi. "A novel technique for multi-class ordinal regression- APDC." Indian Journal of Science and Technology 9.10 (2016): 1-5.
10. Anggoro, Dimas Aryo, and Naqshauliza Devi Kurnia. "Comparison of accuracy level of support vector machine (SVM) and K-nearest neighbors (KNN) algorithms in predicting heart disease." International Journal 8.5 (2020): 1689-1694.
11. Liang, X., Lu, X., & Chen, X. (2020). Deep learning in medical image analysis. Annual review of biomedical engineering, 22, 341-367.
12. Gladence, L. Mary, T. Ravi, and M. Karthi. "Heart disease prediction using Naïve Bayes classifier- sequential pattern mining." the International Journal of Applied Engineering Research ISSN (2014): 0973-4562.
13. Cortes, C., & Vapnik, V. (1995). "Support-vector networks." Machine Learning, 20(3), 273-297.
14. Akkaya, U., Adiguzel, A., & Gulten, A. (2022). Deep learning-based multi-class arrhythmia detection from electrocardiogram signals. Biomedical Signal Processing and Control, 73, 103038.
15. Han, B., Kim, Y. K., Shin, Y. J., Kang, E. Y., & Lee, K. H. (2016). "Automated lung nodule detection in CT images using a deep convolutional neural network." PLOS ONE, 11(3), e0152223.
16. Hussein, M., Ammar, H., & Riad, A. M. (2022). A deep learning approach for heart diseases classification using electrocardiogram signals. Computers in Biology and Medicine, 140, 10486
17. Prasad, Tvs Gowtham, et al. "Cnn Based Pathway Control To Prevent Covid Spread Using Face Mask And Body Temperature Detection." Journal of Pharmaceutical Negative Results (2022): 1374-1381.1911-1917.
18. P. Sai Kiran. "Power aware virtual machine placement in IaaS cloud using discrete firefly algorithm." Applied Nanoscience (2022): 1-9.
19. Peneti, S., Sunil Kumar, M., Kallam, S., Patan, R., Bhaskar, V. and Ramachandran, M., 2021. BDN-GWMNN: internet of things (IoT) enabled secure smart city applications. Wireless Personal Communications, 119(3), pp.2469-2485.
20. Malchi, Sunil Kumar, et al. "A trust-based fuzzy neural network for smart data fusion in internet of things." Computers & Electrical Engineering 89 (2021): 106901.
21. Sangamithra, B., P. Neelima, and M. Sunil Kumar. "A memetic algorithm for multi objective vehicle routing problem with time windows." 2017 IEEE International Conference on Electrical, Instrumentation and Communication Engineering (ICEICE). IEEE, 2017.
22. Sunil Kumar, M., and A. Rama Mohan Reddy. "An Efficient Approach for Evolution of Functional Requirements to Improve the Quality of Software Architecture." Artificial Intelligence and Evolutionary Computations in Engineering Systems. Springer, New Delhi, 2016. 775-792.
23. Kumar, T. P., & Kumar, M. S. (2021). Optimised Levenshtein centroid cross-layer defence for

multi-hop cognitive radio networks. IET Communications, 15(2), 245-256.

24. Natarajan, V. Anantha, et al. "Segmentation of nuclei in histopathology images using fully convolutional deep neural architecture." 2020 International Conference on computing and information technology (ICCIT-1441). IEEE, 2020.

25. Ganesh, D., et al. "Extreme Learning Mechanism for Classification & Prediction of Soil Fertility index." Journal of Pharmaceutical Negative Results (2022): 37-43.

26. D.Ganesh,M.Sunil Kumar., et al. "IMPLEMENTATION OF CONVOLUTIONAL NEURAL NETWORKS FOR DETECTION OF ALZHEIMER’S DISEASE",BioGecko, A Journal for New Zealand Herpetology, Vol 12 Issue 01 2023

27. Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., & Sánchez, C. I. (2017). "A survey on deep learning in medical image analysis." Medical Image Analysis, 42, 60-88.