International Journal of Research in Advanced Engineering and Technology
www.allengineeringjournal.in
ISSN: 2455-0876
Received: 29-05-2025, Accepted: 29-06-2025, Published: 14-07-2025
Volume 11, Issue 2, 2025, Page No. 31-37

A comprehensive review of deep ensemble learning techniques for medical disease diagnosis

Kulkarni Usha Bhimrao[1], **Dr. Sanjay Kumar**[2]
[1] Reseach Scholar, Department of Computer Science, Kalinga University, Raipur, Chhattisgarh, India
[2] Guide, Department of Computer Science, Kalinga University, Raipur, Chhattisgarh, India

Abstract
The fast development of artificial intelligence in healthcare has made a considerable difference in the system of disease diagnosis and prediction. Deep ensemble learning has become one of the powerful techniques that have been able to combine several machine learning and deep learning models to enhance diagnostic accuracy, robustness, and generalization. The present review paper is a thorough analysis of the deep ensemble learning methods used in medical disease diagnosis. It critically compares the traditional classifiers, such as Support Vector Machines (SVM), Naive Bayes and K-Nearest Neighbors (KNN) as well as state of the art deep learning architectures, such as Convolutional Neural Networks (CNNs), ResNet, VGG16 and MobileNetV2. The paper identifies some of the major issues in medical data analysis, including class imbalance, noisy and heterogeneous data, and high dimensionality. The special attention is paid to Synthetic Minority Oversampling Technique (SMOTE) and noise filtering techniques that improve the quality of the data and the performance of the model. Moreover, the effectiveness of ensemble strategies like bagging, boosting, and stacking are discussed in terms of their success in enhancing predictive performance in complex medical data. Results indicate that deep ensemble learning models are more accurate, sensitive and reliable compared to single-model approaches, especially in critical applications, like cancer detection, cardiovascular disease prediction and medical imaging analysis. This review also points out gaps in research and outlines the direction of the research in the future to develop efficient, interpretable and scalable diagnostic systems. In general, deep ensemble learning has great potential as an intelligent clinical-decision-support system to diagnose the disease at an early and accurate stage.

Keywords: Deep ensemble learning, medical disease diagnosis, machine learning, deep learning, healthcare analytics, medical imaging, classification algorithms

Introduction

In 2017, 9.6 million people lost their lives to cancer; the majority of these casualties occurred in less developed nations, as reported by WHO. At this very moment, 422 million individuals throughout the globe are living with diabetes. Every year, it results in 1.5 million deaths. Cardiovascular diseases (CVDs), on the other hand, lead to 17.5 million deaths annually. In 2008 214 360 women in China lost their lives to breast cancer and by 2021, it will be 2.5 million. What such a horrific circumstance originates to patients and their families. Therefore, it is necessary to find the real reasons behind "such a large number of deaths".

According to the WHO, many cancer cases are discovered too late, but with accurate and early identification, over 30% of patients will survive. Thus, to enhance healthcare in our society, we must develop an efficient early illness detection method. Due to their ability to glean useful insights from hierarchical time series clinical data that is large, complex, varied, and subject to frequent usage in medical diagnosis, machine learning (ML) algorithms are becoming more popular. ML approaches may also prevent pathologists and doctors from making medical mistakes due to inexperience, exhaustion, stress, etc. and review medical data quickly and thoroughly.

In previous research, there is a major issue with medical diagnosis as far as classification issues are concerned. There have been several effective methods, such as in the form of neural networks, Naive Bayes, KNN, and SVM. These state of the art classification algorithms did not pay attention to the imbalanced input data, but solely looked at classification accuracy. In case of the imbalance in the input data, the classifier will give more attention to the samples of the dominant class and under emphasise the ones of the minority. Consequently, there will be a higher difficulty in medical diagnosis and reduced classification performances. Due to the aforementioned limitation, we have only considered the binary classification problem and presented a 3 stage ensemble learning scheme of medical diagnosis with imbalanced medical data. SMOTE and cross-validated committees filter are used to resample instances.

SMOTE is superior than under-sampling and may over-represent minority classes by producing fictional instances rather than copying. SMOTE solely synthesised minority, ignoring class noise. Inspired by this weakness, we propose CVCF noise filtering to reduce noisy data and build the combined SMOTE-CVCF data pre-processing technique. Committee-based CVCF noise filter is high-performance. Ensemble learning was introduced to classification in the second phase. Ensemble learning, a relatively new kind of machine learning, can account for the performance of each classifier and the numerous mistakes they may generate, resulting in more accurate classification results than with a single classifier.

The two biggest challenges in ensemble learning for classification are choosing the numerous members of a classifier to create an ensemble and communicating the various classifiers' choices to a coordinated decision. First difficulty is very important. SVM, with its minimal algorithmic complexity and strong resilience to tackle binary classification problems, is the most common and effective classifier. The benefits of categorisation have led to widespread SVM adoption. Past studies concentrated on

SVM classifier parameter optimisation or feature selection, which may be overfitted and provide optimal results. This topic led us to create SVM classification model ensemble members with different diversity architectures. In the last phase, we use weighted fusion to overcome majority voting's disadvantages and evaluate each ensemble member's categorisation contribution. The ideal weight vector of the fusion process is found using hybrid (SAGA).

As far as we are aware, no research has identified clinical disorders using skewed datasets and various diversity patterns of SVM ensemble classifiers. We came up with a new way to use ensemble learning for medical analysis that works around this problem by using data that isn't fair. We believe it may be used as a clinical intelligent diagnostic tool for medical decision makers.

Objectives

To critically review existing literature and analyze the performance of basic classifiers and conventional supervised learning algorithms on medical datasets with respect to their efficiency and accuracy in disease prediction.

Review of Literature

Patel & Shah (2024) [1], We use deep learning to construct a medical identification system to recognise normal states, lung opacity, COVID-19, pneumonia, and other lung illnesses from chest X-rays. The purpose is to enhance diagnostics using ensemble fixed features learning. This will help develop a reliable and affordable diagnostic tool to address the global lung epidemic. The study team utilised COVID-19 pandemic chest radiographs from Kaggle. Enhanced contrast and noise removal are done on raw X-rays. Dataset imbalance is fixed via near-miss resampling. Ensemble learning techniques like two- and three-level methods maximise the capabilities of base learners VGG16, InceptionV3, and MobileNetV2. Model performance is measured by F1-score, accuracy, recall, and precision. User interface and shared web connection are constructed using Python Gradio for remote access. Two-level ensembles using base learner features are classified using a support vector machine. A three-level ensemble's final prediction employs a majority vote method and concatenated features from three machine learning classifiers. The two-stage technique yielded 93% F1 score, recall, accuracy, and precision. The three-level ensemble model excels in recognising COVID-19, pneumonia, lung opacity, and normal states with 94% accuracy. This work uses ensemble learning to enhance lung disease identification from raw chest X-rays, making it unique. Three pretrained networks tuned for autonomous feature extraction replace feature engineers in the model. The approach may help healthcare clinicians, particularly those in low-resource situations, make decisions.

Zhou et.al (2024) [2], Since heart disease (HD) kills people, society, families, and patients pay a significant price. HD diagnosis and treatment are crucial, and real-time prediction may reduce mortality. Deep learning (DL) methods predict HD quicker and better in real time. We compared the merits and disadvantages of DL, ETDL, and integrated DL algorithms, which employ DL with other technologies to anticipate HD. After comprehensive screening, 64 papers from 2018–2023 were selected for exploratory inquiry. A comprehensive literature review of real-time HDP may help future academics understand deep learning and integrated healthcare technologies. The typical datasets used to deploy multiple prediction models are also examined. Furthermore, it reveals any unsolved challenges or difficulties experienced by previous researchers. Lack of large discrete datasets is the biggest obstacle, followed by improving current models.

Mahajan & Moni (2023) [3], Machine learning techniques have enhanced some sickness prediction systems. Ensembles of classifiers perform better and provide more accurate predictions than single classifiers. Despite their widespread use for disease prediction, few studies have compared popular ensemble approaches to well-studied disorders. This research examines five well-studied diseases: diabetes, skin disease, renal disease, liver disease, and heart disease. Finding patterns in ensemble procedures like bagging, boosting, stacking, and voting performance accuracy is the aim. First, we used a straightforward search strategy to discover 45 papers from 2016 to 2023 that employed two or more of the four ensemble methods to treat any of these five illnesses. Stacking (23) has been used less than bagging (41) or boosting (37), but it has been the most accurate (19 out of 23). Voting was the second-best ensemble approach in this assessment. For diabetic and skin condition papers, stacking consistently performed better. Bagging worked better five times for renal sickness, although boosting performed better four times for liver and diabetes. Stacking is the most accurate sickness prediction method among four. Our investigation reveals that ensemble approaches perform differently on common disease datasets. This study's findings will assist academics choose the best ensemble model for predictive disease analytics and understand the latest trends and hotspots in ensemble learning-based illness prediction models. This paper also compares ensemble approaches to regularly used sickness datasets.

Fayez, Mustafa & Kurnaz, Sefer. (2021) [4] ML is also considered a cutting-edge method, thus it's best left to experts. A large number of medical and biological researchers are unable to conduct their experiments using this device. The purpose of this research is to disprove this outdated idea that has been prevalent. We claim that the emergence of new high-performance machine learning methodologies has made it possible for researchers in the field of biomedicine to construct competitive machine learning models in a short amount of time with little to no comprehension of the mechanisms that underlie them. Python, the most advanced programming language currently accessible, serves as the foundation for both of the components that make up this cutting-edge system. Neighbourhood Cleaning Rule (NCL) implementation is the first stage in starting feature engineering and preprocessing. NCL is a high-performance resampling approach. Second, there are machine learning models that are considered to be state-of-the-art and have great performance, such as XGBoost and AutoML, in addition to ensemble bagging models. Last but not least, we are certain that our advances will open the way for wider clinical application of AI techniques and improve the capacity of medical professionals to comprehend machine learning with the assistance of high-performance machine learning technology.

Akinbo, R. S., *et al*. (2021) [5] Illness categorisation utilising machine learning methodologies is now being used as a precise treatment for scientific improvement. As a result of the meteoric rise of machine learning systems, medical imaging research has mushroomed. Medical image classification using machine learning techniques is the focus of this research. Among the many appealing applications of machine learning is the possibility of automated medical image tagging, which might improve prediction and accuracy. We will have a better understanding of illnesses and be able to tackle some of the most intractable challenges thanks to advancements in medical image processing technologies. This paper compares and contrasts the key ensemble learning techniques, analyzes the conceptual framework, use, and constraints of machine learning and deep learning. It not only analyses and reviews previous research, but also looks at the most cutting-edge work on ensemble learning. The research covers a very diverse range of medical image formats, their applications, benefits and uses. We proposed that a system of medical picture classification based on ensemble machine learning would enhance the accuracy and the performance of the system. The incorporation of emerging technology in medical images can positively impact analysis, prediction, classification, and early detection. The end result will be a higher quality of life to people and more therapeutically useful machine learning applications.

Gupta *et al*. (2021) [6] provides an explanation of the image registration survey. In this study, many sorts of algorithms were used to perform anatomical and functional medical picture registration. There are many various kinds of medical pictures, such as the MRI image, the CT scan, the PET scan, the FLAIR scan, and so on. The purpose of this is to examine the image registration methods that may be processed between the functional medical images and the anatomical medical images. The results of a study comparing several image registration methods showed that the supervised learning model might get better at predicting features between images.

Kesav *et al*. (2021) [7] A texture feature learning model was shown as a way to sort brain MRI images into different groupsThe image's feature values were extracted from the background pattern using the S-transform. MR images must be preprocessed & bit-mapped using fuzzy-C means clustering. The goal was to locate the problematic area of the brain picture. This model uses the discrete orthonormal S-transform to extract picture characteristics. A PCA + LDA combo may then reduce the feature array's dimension. The (ADBRF) classifier grouped this. Random forest was used as the primary predictor to extract the image.

Hassan Habib *et al*. (2021 [8]) proposed a combination of image segmentation, classification, and feature extraction to investigate the proliferation of aberrant brain cells. In order to locate the tumour spot and image cluster in the image matrix, the watershed algorithm and the threshold segmentation technique were used during the image segmentation stage. To show image features, the image FE mixed model employs a multitude of techniques, such as MSER, FAST, and Harlick. The next step was to utilise many models to create good assumptions about what the labels should be on the brain MRI pictures. The pattern learning model deconstructs the image into its component parts before constructing the categorisation model.

Das *et al*. (2021) [9] suggested a group model for an image classification that could use the X-ray image collection to separate normal patients from COVID-19 cases. The goal is to look at how well the ensemble model of the picture analyser works. CNN, RNN, LSTM, and Gated Recurrent Units were some of the classifier models that were used to make the training set and data. So that the training could go as smoothly as possible, the training tasks were carried out at the same time. The fuzzy min-max prediction was used to get rid of classification function error and make the final choice. To make sure the classification was correct, this was also checked on a picture of the brain.

Rezaei *et al*. (2020) and Zhu *et al*. (2021) [10, 14] Use the ensemble training model to put MR brain pictures into groups, and suggest a weighted vote method. The hurt part of the brain could be found using SVM-based picture segmentation on MR slices. In feature learning, 42 features are taken out and put together in different ways to make a feature vector. Differential Evolution (DE) is used to get the eleven most important traits in this case. The (KNN), (WSVM), and histogram overlap (HIK-SVM) algorithms are used to sort and predict the tumours in the MR slices based on the chosen features. Several classification models were put together, and an ensemble method based on (MODE) helped them work.

Machine Learning in Medical Diagnosis

Research from 2020 [10] by Dlamini *et al*. indicates that machine learning's capacity to predict healthcare issues is making it a very important technology. This chapter gives an outline of DL algorithms and ML, and it covers subjects such as foetal heart disease, pneumonia, cardiovascular sickness, breast cancer, and other similar conditions. The chapter goes on to talk about how these algorithms may be used in applications that evaluate disease. Medical records, x-rays, genetic data, and demographic information are only some of the complex clinical data that machine learning algorithms can analyse. In this way, potential dangers and hidden relationships may be discovered, even if human specialists fail to see them. Johnson *et al*. (2021) [14] say that combining clinical knowledge and data-driven insights could completely change how healthcare decisions are made, how treatments are planned, and how illnesses are diagnosed early.

The ML approach gives computers new abilities and improves their performance over time. Machine learning does this without explicit programming. Data-driven knowledge uses algorithms to find patterns, correlations, and facts for analysis and decision-making. Data analysis for more accurate and updated patient health estimations in healthcare. Some of ML's applications include data analysis, finance, healthcare, recommender systems, autonomous machines, language processing, image processing, and an endless list of others. It has also improved game-changing innovation that affects business, academics, and society.

Deep Learning Approaches in Healthcare

In this part, a comprehensive explanation of the DL architectures that are used in the detection of pneumonia is presented. CNNs, Inception, ResNet, VGG16, MobileNetV2, DenseNet121, and RNNs among others can be considered to identify their nuances and effectiveness. This exploration was useful in the development of the fields.

1. **CNNs:** CNNs are Convolutional Neural Networks. They are DL models used for visual identification like object recognition and photo sorting. CNNs feature non-linear activation functions which find deep

connections like ReLU. Although their interpretability & overfitting may be addressed by transfer learning, they are best for visual object identification, categorisation, and face recognition. CNNs' convolutional layers can learn systematically, making them effective at extracting attributes. CNNs specialise in video data. Black box models are challenging to evaluate, and massive data sets may induce overfitting (Edureka, 2018).

2. **Inception:** The computing power of the Inception neural network architecture, which is specifically designed to operate with images has been greatly praised. Two distinguishing features are the global average pooling that gets rid of superfluous parameters and the Inception modules that handle multi-scale feature collection. It is accurate in the working but more complex in nature and thus could require more time to train. Picture categorisation and contests like the ImageNet Challenge are only a few application areas of Inception. Inception fits best in a job where resources available are vital in terms of their efficiency. The strengths of the initiation modules in the network are not a secret. The network is also in a position to gather data of varying size by use of such modules hence resulting in better performance and efficiency. It is possible that the training process will be more complex because of the more complex model (Fischer and Igel, 2012) [15].

3. **Residual Neural Network (ResNet50):** There are a lot of problems with training deep neural networks, but the new ResNet design can fix them all. The vanishing gradient issue, eliminating the remaining blocks, and eliminating skip links to allow free flow of information are some of the key issues that should be addressed. Identifying objects, splitting up pictures, and putting pictures into groups are all things that ResNet could be useful for. Also, it simplifies the training of deep neural networks. It had a significant impact on the field, and the designs of neural networks that followed it were modified due to it. Addressing the issue of disappearing gradients in deep learning presents a compelling challenge. Negative aspects include the possibility that a more complex model may make interpretation more challenging (Hinton & Salakhutdinov, 2006) [16].

4. **Visual Geometry Group (VGG16):** With regard to the classification of photographs, VGG-16 is well-known for being both easy and effective. The process of feature extraction is made more efficient by the use of repeated blocks of small-sized filters; nevertheless, there are a few downsides, including the possibility of overfitting and an increase in the amount of processing that is required. Regardless of these shortcomings, VGG16 continues to be a popular choice for benchmarking and baseline research in applications that are connected to images. When it comes to feature extraction, the architecture is made more straightforward and effective by the use of repetitive blocks of very small convolutional filters. According to Venugopalan *et al*. (2021) [20], the disadvantages of this model include a huge number of parameters, which makes it computationally costly and prone to overfitting.

5. **MobileNetV2:** The efficient, low-cost, and diminutive CNN MobileNetV2 is an excellent pick for mobile and edge devices because to its focus on efficiency, reduced computational cost, and smaller memory footprint. The perfect fit for creating programs for mobile devices, thanks to its adaptable design that finds a happy medium between performance and resource constraints. This feature provides benefits, such as a smaller memory footprint and lower computational costs, and is designed for use on mobile and edge devices.The disadvantages include a reduction in accuracy when compared to larger models, especially when it comes to more challenging tasks (Sivaranjini & Sujatha, 2019) [21].

The report emphasised the impact that these methods have had in several fields, such as genomics, medical imaging, e-health, medicine development, and disease prognosis. On top of that, it has shown the usefulness of CNNs, RNNs, and generative adversarial networks in medical settings. These facilities have significantly contributed to the drug development processes, enabled more accurate analysis of medical images and provided a more accurate risk classification of patients. With an eye toward the future, current research should concentrate on resolving outstanding issues and improving deep learning methods for easier integration into clinical practice. Lastly, the future of healthcare is bright owing to the application of to deep learning techniques.

Ensemble Learning Techniques

The goal of using ensemble ML techniques is to increase the system's prediction performance by integrating many models. Ensemble machine learning techniques have been more successful than single-model methods, when it comes to predictive modelling and data analysis. These strategies have great benefits when compared to the more traditional methods. This research aims at digging deep into the results of a thorough literature review that examines the feasibility and performance of ensemble techniques, i.e. bagging, boosting and stacking. These papers highlight the use of several learning models in tandem to improve prediction accuracy.

For model stability and variance reduction, ensemble approaches like Bagging on Random Forests may be helpful. In the meanwhile, AdaBoost and Gradient Boosting show that boosting may minimise bias, especially when employed on less proficient learners. Researchers also study stacking, which goes beyond ensemble approaches and integrates the predictions of many models trained individually.

Comparative examination of various approaches may assist explain their differences, taking into consideration over-fitting, under-fitting, and model dependencies. The studies also emphasise feature importance analysis, which identifies crucial elements that might greatly affect prediction accuracy. Traffic accident forecast depends on collision type, weather, and road surface. Intrusion detection systems also use relevant variables to determine model performance. This literature study shows that ensemble techniques are practical and have many applications. Most datasets' fundamental flaws have been addressed by these approaches, improving predictions.

The present effort to identify future research domains intends to improve prediction accuracy, reduce computing burden, and apply ensemble techniques in non-traditional ML paradigms. This study will help us understand ensemble approaches and their potential use in the ever-changing machine learning industry. An ensemble of models has two benefits over a single model. List of them follows. Performance: An ensemble may outperform and predict more accurately than any single model. Robustness in noisy datasets is an ensemble's capacity to remain stable and generalise under various situations.

Deep Ensemble Learning For Disease Diagnosis

When it comes to healthcare systems, Ensemble DL is superior to both conventional ensemble learning and classic DL. Improved accuracy, robustness, and generalisability are the outcomes of merging predictions from several classifiers. When working with healthcare datasets that are varied, unbalanced, and noisy, this becomes much more apparent. This technique outperforms standard ensemble learning when it comes to enhancing sensitivity and specificity and lowering the danger of overfitting in critically important diagnostic applications.

1. **Medical Imaging**

   Ensemble deep learning is an effective computer vision method. Computer vision employs video and pictures frames by default for prediction and categorisation. Medical imaging is used using computer vision algorithms to enhance diagnosis and therapy. It covers radiology, ophthalmology, pathology, and dermatology. CT, PET, and MRI generate such medical pictures. They will promote early medical care and diagnosis. Seven standard medical imaging methods exist.

2. **Disease diagnosis and prevention**

   The last several years have witnessed an upsurge in global health crises, highlighting the need to detect and prevent sickness early. Another important part of illness diagnosis is determining the underlying medical condition that caused symptoms. Ensemble deep learning may assist since it uses previously taught data to forecast illnesses earlier. Big datasets have improved prediction accuracy, overfitting, and generalisation in a wide range of medical tasks, including cardiovascular diseases, diabetic retinopathy, pneumonia, as well as others, making this practice widespread. Ensemble models reduce mortality owing to early detection and therapy and are more successful than single-model techniques. Use feature extraction, good preprocessing, and transfer learning.

Challenges in Medical Data Analysis

As a result of its ability to facilitate early illness detection, personalised therapy, and better patient outcomes, medical data analysis has emerged as an essential component in contemporary healthcare systems. An vast amount of data pertaining to healthcare is being continually created as a result of the fast rise of technology such as biomedical imaging, wearable devices, and digital health records. Nevertheless, despite its promise, medical data analysis is confronted with a number of severe problems that restrict the efficiency and dependability of prediction models.

- **Heterogeneity of Medical Data:** When analysing medical data, one of the most common issues is dealing with the many data sources. (EHRs), the results of lab tests, imaging systems, personal sensors, and professional notes are just some of the places where medical data is gathered. There are differences in the format, structure, and semantics of these data sources. Data that is organised, such as numerical lab findings, coexists alongside data that is not structured, such as notes from a physician, and data that is semi-structured, such as medical reports. Because of this variety, it is becoming more challenging to efficiently integrate and analyse the data. It takes a lot of effort to preprocess and change data so that typical machine learning models can use it, and having data in a consistent format is generally required. Interoperability across different healthcare systems is further complicated by the absence of standardised data formats, which also makes it difficult to construct generalised predictive models on a large scale.

- **Missing and Incomplete Data:** The majority of the time, medical datasets are missing or partial information. The probable reasons are human error during the data entry, variability in the diagnostic processes or the unwillingness of the patients to comply. Data loss can greatly decrease the effectiveness of machine learning models and give unreliable or prejudiced predictions. The handling of missing data is a challenging problem since even simple methods, such as mean imputation, may not be able to maintain the distribution of the data that is being used. In the case of the integrity of data upkeep, more sophisticated imputation methods are frequently required; however, they make the calculation more complicated.

- **Noisy and Inconsistent Data:** The fact that medical data is noisy and inconsistent is because of many factors such as changing measuring instruments, human errors as well as differences in clinical procedures. Data noise may hide key patterns and lower prediction model accuracy. Data noise may sometimes be quite distracting. Irrelevant discrepancies may also come from lab test findings in different laboratories. Duplicated entries, contradictory records, and poor data input also degrade data quality. These difficulties need high-level data cleaning & preprocessing procedures that require subject expertise and complicated algorithms.

- **High Dimensionality and Irrelevant Features:** Medical databases may be cluttered with extraneous data and duplication. The "curse of dimensionality" occurs when overfitting and complexity reduce machine learning model performance. This may be due to high-dimensional data. Use feature selection or dimensionality reduction to find the most important characteristics. However, due to the quantity of nonlinear interactions and the intricacy of variable relationships, medical data may not give significant characteristics. Incorrectly selecting attributes may lose diagnostic information.

- **Class Imbalance Problem:** This discrepancy makes machine learning algorithms susceptible to majority-class bias, resulting in poor minority class categorisation. Traditional evaluation methods like correctness might be deceptive in such cases. Hence, resampling, cost-sensitive learning, and sophisticated assessment measures (F1-score, accuracy, and recall) are needed to appropriately cope with class imbalance.

Conclusion

In this review article, a comprehensive introduction to deep methods of ensemble learning and their growing relevance in diagnosis of medical illnesses is provided. Although there is still a use of classical ML models, the research demonstrates that they frequently have issues with class imbalance, noisy datasets, and poor generalisability among others. Combining multiple model architectures and decision-making processes, deep learning models, particularly when applied in ensemble techniques, are better-performing. Preprocessing methods such as SMOTE, noise filtering methods need to be incorporated to fix data imbalance and enhance the quality of medical datasets. Ensemble methods like bagging, boosting, and stacking build upon the accuracy and resiliency of the predictions made by many classifiers. Moreover, deep ensemble models to medical imaging and disease prediction tasks have demonstrated exceptional improvements in this field. Still, even with recent advancements, computerised complexity, non-interpretability, and issues with various and inadequate medical data are still considered to be a major obstacle. Lightweight, explainable, and scalable ensemble models that are easy to incorporate into real-world clinical settings should be the focus of future research. Deep learning can be significantly enhanced by techniques that blend deep learning with more traditional methodologies and domain expertise, which can make a significant contribution to the diagnostic performance. Deep ensemble learning is an excellent choice to make, as far as smart healthcare systems are concerned. Its forecasts of early illness can be of great benefit to clinicians and healthcare professionals as the forecasts are accurate and dependable. Such a tool will ultimately result in improved patient care and more effective medical decision-making.

References

1. Patel M, Shah M. Deep ensemble learning model for diagnosis of lung diseases from chest X-ray images. Indian Journal of Science and Technology,2024:17(8):702–712. https://doi.org/10.17485/IJST/v17i8.3151
2. Zhou C, Dai P, Hou A, Zhang Z. A comprehensive review of deep learning-based models for heart disease prediction. Artificial Intelligence Review, 2024, 57(10). https://doi.org/10.1007/s10462-024-10899-9
3. Mahajan P, Uddin S, Hajati F, Moni MA. Ensemble learning for disease prediction: A review. Healthcare,2023:11(12):1808. https://doi.org/10.3390/healthcare11121808
4. Fayez M, Kurnaz S. Novel method for diagnosis diseases using advanced high-performance machine learning system. Applied Nanoscience, 2021, 13. 10.1007/s13204-021-01990-6.
5. Akinbo RS, Daramola OA. Ensemble Machine Learning Algorithms for Prediction and Classification of Medical Images. In Artificial Intelligence. IntechOpen, 2021. https://doi.org/10.5772/intechopen.100602
6. Gupta S, Gupta P, Verma VS. Study on anatomical and functional medical image registration methods. Neurocomputing,2021:452(10):534–548.
7. Kesav OH, Rajini GK. Automated detection system for texture feature-based classification on different image datasets using S-transform. International Journal of Speech Technology,2021:24(2):251–258.
8. Habib H, Amin R, Ahmed B. Hybrid algorithms for brain tumor segmentation, classification and feature extraction. Journal of Ambient Intelligence and Humanized Computing,2021:1(1):1–22.
9. Das A, Mohapatra SK, Mohanty MN. Design of deep ensemble classifier with fuzzy decision method for biomedical image classification. Applied Soft Computing,2021:115(1):305–313.
10. Rezaei K, Agahi H, Mahmoodzadeh A. A-weighted voting classifiers ensemble for the brain tumors classification in MR images. IETE Journal of Research,2020:1(1):1–14.
11. Mohebbian M, Walia E, Habibullah M, Stapleton S, Wahid KA. Classifying MRI motion severity using a stacked ensemble approach. Magnetic Resonance Imaging,2021:75(1):107–115.
12. Subash Chandra Bose S, Sivanandam N, Sundar PV. Design of ensemble classifier using statistical gradient and dynamic weight LogitBoost for malicious tumor detection. Journal of Ambient Intelligence and Humanized Computing,2021:12(6):6713–6723.
13. Dlamini Z, Francies FZ, Hull R, Marima R. Artificial intelligence (AI) and big data in cancer and precision oncology. Computational and Structural Biotechnology Journal,2020:18:2300–2311. https://doi.org/10.1016/j.csbj.2020.08.019
14. Johnson KW, Torres Soto J, Glicksberg BS, Shameer K, Miotto R, Ali M, *et al*. Artificial intelligence in cardiology. Journal of the American College of Cardiology,2021:71(23):2668–2679. https://doi.org/10.1016/j.jacc.2018.03.521
15. Fischer A, Igel C. An introduction to restricted Boltzmann machines. Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications, 2012, 14–36. https://doi.org/10.1007/978-3-642-33275-3_2
16. Hinton GE, Salakhutdinov RR. Reducing the dimensionality of data with neural networks. Science,2006:313(5786):504–507. https://doi.org/10.1126/science.1127647
17. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. International Conference on Learning Representations (ICLR), 2015.
18. Sandler M, Howard A, Zhu M, Zhmoginov A, Chen LC. MobileNetV2: Inverted residuals and linear bottlenecks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018, 4510–4520.
19. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, 770–778.
20. Venugopalan S, Tong L, Hassanzadeh HR, Wang MD. Multimodal deep learning models for early detection of Alzheimer's disease stage. Scientific Reports,2021:11(1):3254. https://doi.org/10.1038/s41598-021-82031-2

21. Sivaranjini S, Sujatha CM. Deep learning-based diagnosis of disease in plants using MobileNet. International Journal of Engineering and Advanced Technology,2019:8(6):193–198.
22. Dietterich TG. Ensemble methods in machine learning. International Workshop on Multiple Classifier Systems, 2000, 1–15. https://doi.org/10.1007/3-540-45014-9_1
23. Breiman L. Random forests. Machine Learning,2001:45(1):5–32. https://doi.org/10.1023/A:1010933404324
24. Friedman JH. Greedy function approximation: A gradient boosting machine. Annals of Statistics,2001:29(5):1189–1232.
25. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: Synthetic minority over-sampling technique. Journal of Artificial Intelligence Research,2002:16:321–357.
26. Goodfellow I, Bengio Y, Courville A. Deep learning. MIT Press, 2016.