STM JOURNALS

Research & Reviews: A Journal of
Medical Science and Technology

https://journals.stmjournals.com/rrjomst/

ISSN: 2319-3417 (Online)
ISSN: 2349-1272 (Print)
Volume 14, Issue 3, 2025
September–December
DOI(Journal): 10.37591/RRJoMST

Research RRJoMST

Early Disease Detection Using Artificial Intelligence

Pavandeep Kaur[1], Parul Gogia[1,*], Aekansh Khandelwal[1], Rishabh Raj[1]

Abstract

Growth in artificial intelligence and machine learning now makes it possible for the healthcare sector to be totally transformed by a new chapter, particularly in the era of medical image analysis. This study focuses on harnessing these advancements to develop a sophisticated model for early disease detection across diverse medical domains, primarily in skin disease. By integrating diverse datasets and leveraging advanced algorithms, our methodology aims to identify subtle disease indicators at their inception, facilitating timely interventions and personalized treatment strategies. Through meticulous data collection, preprocessing, and exploratory analysis, the study establishes the groundwork for the development of robust AI models capable of interpreting complex medical imaging data. The proposed methodology emphasizes the integration of domain-specific clinical expertise to ensure the clinical relevance and interpretability of the models. Rigorous validation and evaluation demonstrate the efficacy and generalization capacity of our approach, paving the way for its seamless integration into clinical practice.

Keywords: Early detection, artificial intelligence, machine learning, medical imaging, disease prediction

INTRODUCTION

In the era of healthcare industries, early detection of a disease plays an important part in the life of a patient, and when doctors detect the disease in the initial stage, it saves cost, patient health, and ultimately saves the patient's life going to a severe condition. The ability to identify the disease during its early stage, such as skin cancer detection and infectious diseases in their very early stages, challenge healthcare professionals. Although there has not been very rapid progress at large, strides that are being made in artificial intelligence (AI) and Machine Learning (ML) hold promise for their revolutionary impact in disease detection and diagnosis, such as the retina-eye-scan, used for detecting retinal diseases.

A model or advanced approach is recommended. Despite the limited progress in broader areas, advancements in artificial intelligence (AI) and machine learning are expected to transform the detection and diagnosis of diseases Bottleneck [1]. For instance, the retinal eye scan is one such technology that can detect eye diseases that are potentially sight-threatening, their genetic records of data, and past clinical records by identifying the patterns about disease risk and best diagnosis at their initial stage itself to prevent them from being conversion of a small disease into a larger one. By the capabilities and the power of AI and ML, this system aims to provide patients by providing timely and accurate reports of their health.

***Author for Correspondence**
Parul Gogia
E-mail: parulgogia4550@gmail.com

[1]Student, Department of Computer Science and Engineering, Apex Institute of Technology, Chandigarh University, Mohali, Punjab, India

Received Date: July 10, 2025
Accepted Date: August 19, 2025
Published Date: September 30, 2025

Citation: Pavandeep Kaur, Parul Gogia, Aekansh Khandelwal, Rishabh Raj. Early Disease Detection Using Artificial Intelligence. Research & Reviews: A Journal of Medical Science and Technology. 2025; 14(3): 11–19p.

The Significance of Early Disease Detection

The importance of early disease detection cannot be overstated. Early identification of conditions, like cancer and tuberculosis, can dramatically increase survival rates by catching the disease before it advances. It detects the disease at its rising stage when the illness rate is very low. Thereby

© STM Journals 2025. All Rights Reserved

maximizing the effect of treatment. It ensures the provider initiates treatment at an early stage. For example, in the case of cancer patients, early detection and diagnosis increase long-term survival and improve the chances of successful treatment. As tumors are more localized and smaller, they need less aggressive treatment. Secondly, early disease detection can prevent the disease from advancing to more advanced tissues. Early knowing of the disease slows down the disease, preserving patients' quality of life, reducing the costly medication in the future.

Early disease detection also helps in modernizing the healthcare industry and combines resources and expenditure. By identification, the healthcare system prioritizes preventive measures and the burden associated with long hospital stays and long-term treatment and fosters a more sustainable and equitable healthcare ecosystem.

Challenges in Early Disease Detection

Although early disease detection is widely recognized as beneficial, several challenges hinder its effectiveness and overall adoption. Heterogeneous complexity and variability in disease presentation among different patient populations represent one of the main challenges. Diseases tend to have heterogeneous presentations that are difficult to detect and to diagnose precisely, especially in their early phases when the symptoms are usually subtle or even non-specific.

Moreover, medical imaging data interpretation is specialized and prone to inherent human errors and biases. Even veteran radiologists could struggle with identifying slight anomalies or separating benign lesions from malignant lesions, thereby creating diagnostic uncertainty and variability in practice.

Additionally, the high quantity and complexity of medical imaging data in medical environments presents logistical challenges to manual analysis and interpretation [2, 3]. Standard image analysis methodologies are based predominantly on subjective visual evaluation, which is time-consuming, vulnerable, and labor-intensive to interobserver variation.

The Promise of AI and Machine Learning

With the assistance of magnetic resonance imaging (MRI) for medical imaging data sets will be examined by other AI software. Artificial intelligence systems are capable of detecting illness at an early stage with the aid of information retrieved from computed tomography (CT) and MRI scans. All the MRIs were read by examining the presence or absence of acute stroke [4]. It will handle enormous amounts of medical picture data utilizing artificial intelligence methods.

In addition, AI-enabled mechanisms can enhance human expertise and allow medical professionals to care for complex clinical images more thoroughly and effectively. Each stage of the course can serve as a decision support mechanism, thereby helping sort cases, identify suspicious findings for further review, and offer quantitative severity and progress measures for disease conditions.

RESEARCH OBJECTIVES

The best AI system for early disease diagnosis is the principal objective. And the most sensible approach is to take advantage of the power of artificial intelligence and the existing machine learning capabilities within medical imaging technologies. Specifically, the new system will focus on detecting early symptoms of serious illnesses like cancer, tuberculosis, and other neurological disorders using X-rays, MRIs, and other medical images.

By harnessing cutting-edge technologies, AI and ML, this technology aims to enhance the scalability, accuracy, and efficiency of the detection and diagnosis of disease, and ultimately improve patient outcomes and healthcare delivery. We undertook extensive testing to validate that the AI we have developed for deployment in health care would operate just as well under real-life operating conditions. The development process of the AI system will be discussed in more detail in the following sections, which include the methods for data collection, preprocessing, model development, and evaluation. We

© STM Journals 2025. All Rights Reserved

will present the results of several experiments used in assessing the functionality of the system, and we will discuss the possible implications of our results on future clinical trials and research projects in real-world medical environments.

LITERATURE SURVEY

Table 1 presents the most significant results from several studies on applying AI methods for melanoma and skin cancer early diagnosis. The table includes the authors, disease to be targeted, AI technique used, the ML algorithm utilized for each case rest of them and obtained accuracy rates. Deep learning methods are predominant in the identification of diseases, and according to a chart, all charts notably use convolutional neural networks (CNNs), like Inception-v3, ResNet, and DenseNet, to obtain an accuracy of more than 90%.

Table 1. Summary of reviewed papers and applied techniques in early disease detection using AI.

Reference	Authors	Disease	AI Technique	Machine Learning Algorithm	Accuracy (%)
[5]	Esteva, Kuprel, et al.	Melanoma	Deep Learning	Inception-v3 (CNN)	96.2
[6]	Brinker, Hekler, et al.	Melanoma	Deep Learning	ResNet (CNN)	94.8
[7]	Zahra Waheed; Amna Waheed; et al.	Skin Cancer	Machine Learning	ResNet (CNN)	96
[8]	Haenssle, Fink, et al.	Melanoma	Deep Learning	AlexNet (CNN)	95.5
[9]	Diwakar Gautam, et al.	Melanoma	Machine Learning	Random Forest	79.81
[10]	Fujisawa, Otomo, et al.	Melanoma	Deep Learning	Inception-ResNet (CNN)	93.7
[11]	Kousis, Perikos, et al.	Skin Cancer	Machine Learning	CNN	92.25
[12]	Codella, Gutman, et al.	Skin Cancer	Deep Learning	U-Net (CNN)	94.2
[13]	Esteva, Thrun, et al.	Melanoma	Deep Learning	DenseNet (CNN)	96.4
[14]	Han, Kim, et al.	Melanoma	Machine Learning	Gradient Boosting	92.9
[15]	Brinker, Hekler, et al.	Melanoma	Deep Learning	VGG16 (CNN)	95.1
[16]	Mohamed A. Kassem, Khalid M. Hosny et al.	Skin Cancer	Convolutional Neural Networks	Inception-v3 (CNN)	93.31
[17]	Codella, Gutman, et al.	Skin Cancer	Artificial Intelligence	Support Vector Machine (SVM)	90.7
[18]	Evgin Göçeri	Melanoma	Deep Learning	MobileNet (CNN)	91.33
[19]	Brinker, Hekler, et al.	Skin Cancer	Machine Learning	CNN	90.8

METHODOLOGY

The datasets are drawn from various sources Kaggle, the ACM digital library, Semantic Scholar, and the IEEE Xplore Library. The source has diverse information, from disease symptoms to patient demographics and physical environment found in real-world situations. We have preprocessed the data and applied a CNN algorithm that detects from images and analyzes to match the pre-trained datasets.

Understanding the Problem

This research will focus on the ever-growing role of medical imaging technology in the early detection of disease. It acknowledges how proactive disease detection can play a role in ultimately enhancing the health outcome for both patients and health systems. It is centered on how more sophisticated computational methods to be applied to current methods of disease detection utilizing medical imaging. The research addresses a myriad of diseases in a series of medical domains, including oncology, neurology, cardiology, and radiology. Each disease will bring with it different considerations and challenges in detecting the disease at an early stage (e.g., the complexity of imaging data, initial signs of disease may be imperceptible, and variability between patients).

© STM Journals 2025. All Rights Reserved

Collect Data

The data collection process is pivotal for the success of the research, providing the foundation upon which the subsequent analyses and methodologies are built. The following steps outline the methodology for data collection:

1. Locate and choose datasets with medical imaging data tied to the diseases in your show. When selecting datasets, you should evaluate:
    a. *Variety:* Include variety in imaging modalities (e.g., MRIs, CT scans) [4], types of diseases, and variability in patients, thus encompassing variation in the manifestation of disease.
    b. *Size: Choose datasets that provide enough sample sizes to effectively run model-training/transformation/creation and evaluations.*
    c. *Annotation: Ensure that the imaging data has clear disease labels/annotations.*
2 *Data Acquisition:* Obtain the selected dataset from reliable sources such as medical research repositories, hospitals, or collaborative research organizations. Ensure you confirm with the data protection regulations, like HIPAA, and have any required ethical approvals if required.

Clean and Process Data

The initial critical components of preparing the medical imaging data collected for analysis are data cleaning and data preprocessing.

This entails converting raw data into a structured, standardized format that can be used for the modeling and analysis phase [5]. The methods for cleaning and processing data include the following:

1. Noise Reduction and Artifact Removal
    a. The imaging data requires identification along with correction of noise and artifacts, and condition bits and inconsistencies that occur due to limitations during acquisition or production.
    b. Signal-to-noise ratio improvement, together with imaging data quality optimization, requires the implementation of noise reduction techniques like Gaussian filtering and median filtering.
    c. The application of artifact removal standards helps correct acquisition-induced image distortions and anomalies.
2 Image Registration and Alignment:
    a. Image registration and alignment methods involve establishing some consistent spatial position across the various modalities or snapshots in time.
    b. Transformation algorithms can help to register the images by aligning the spatial position of anatomical landmarks and fiducial markers.
    c. In addition to improving the correction of distortions and motion artifacts, spatial alignment has improved the level of precision available for feature extraction and analysis.
3 Intensity Normalization and Standardization
    a. Normalizing pixel intensities to lessen differences in imaging parameters (e.g., scanner settings, acquisition protocols, and use of contrast agents).
    b. Normalizing image intensities across the data set to reduce the impact of imaging hardware or protocols.
    c. Normalizing images to the same scale or reference frame for consistent extraction of features and modeling.
4 Feature Extraction and Selection
    a. Employing image processing methods such as texture analysis, to derive useful information from pre-processed images.
    b. Choose informative features based on their discriminative power, relevance to disease pathology, and computational efficiency.
    c. Use dimensional analysis techniques to reduce complications of feature space while maintaining data of attributes.
5 Data Augmentation and Synthesis
    a. The models require additional data enhancement, which simultaneously improves their strength and produces unique content.

© STM Journals 2025. All Rights Reserved

   b. Perform image enhancement through rotation, flipping, and elastic deformation methods to increase their size.
   c. Use generative models, such as generative adversarial networks, to create synthetic images that increase the dataset and replicate different disease presentation patterns.
6 Data Splitting and Cross-Validation
   a. Researchers separate the preprocessed dataset into separate training and testing subsets that allow for the evaluation of model performance.
   b. Stratified sampling is also used to achieve an equal representation of all classes and demographic groups across the datasets. Cross-validation methods provide a good way to assess the model's generalization and stop the model from overfitting.

The research team prepares medical imaging data by following systematic cleaning and processing steps to have a suitable dataset to further develop the model that detects diseases early.

Exploratory Data Analysis

The first step of medical imaging data analysis to detect diseases in their early stages involves conducting exploratory data analysis (EDA). Medical researchers start by gathering fundamental data about the dataset that includes sample quantity, along with the complete set of imaging modalities and all disease categories. The following step involves computing summary statistics, which reveal the distributional patterns of important imaging features, including average values, median values, standard deviation, and value span. Researchers examine medical images or image slices through visual representations to better understand how diseases present themselves and how anatomical structures appear, together with potential imaging errors. Researchers examine the way class labels are distributed throughout the dataset to determine whether the classes are balanced or imbalanced using visual two-dimensional representations such as bar charts and violin plots.

Exploring how class labels relate to other demographic variables can assist us in determining if there are potential hidden variables affecting the findings. Methods for dimension reduction allow data to be displayed, while enabling the user to observe natural clustering behaviors and examine the effectiveness of features. Detection methods will identify fake outliers that may arise from poor data quality. Temporal analysis allows for research that generates additional detail for disease progression and provides utility for treatment outcomes and modifications for imaging biomarkers over time. Engaging in the appropriate exploration of data will provide the researcher with valuable experiences useful in model development and hypothesis testing, focusing on early disease detection via analysis of medical images.

RESULT ANALYSIS AND VALIDATION

To assess the proposed disease prediction model, four performance evaluation metrics are implemented [7]. The confusion matrix includes true positives (TP) when a target is correctly predicted as a chronic patient, true negatives (TN) when persons without illnesses are accurately predicted, false positives (FP) when a healthy individual is erroneously predicted to be sick, and false negatives (FN).

a. *Accuracy:* The ratio of correctly predicted values to all predicted values is known as classification accuracy.
b. *Precision:* Precision or Positive predictive value (PPV) refers to the fraction of accurate predictions, all true and all false predictions.
c. *Recall:* The following represent a mathematical formula of recall, sensitivity, or true positive rates (TPR), the ratio of true positive to total true positive.
d. *F1-Score:* The F-measure (Fβ), obtained from measuring precision and recall, is the weighted average of these values. F1-score is more important than accuracy in the presence of an imbalanced class distribution. Besides which it is useful in differentiating between false positive or negative values where they are not equal.

To calculate all the precision, accuracy, and recall, firstly ensure to build a CNN model.

© STM Journals 2025. All Rights Reserved

Building CNN Model (Figures 1–4)

1. *Convolutional Layer:* Filters/Feature maps that are used to transform images. This is called the convolutional layer.
2. *Pooling Layer:* Max pooling is useful for down-sampling. It reduces computational costs and also reduces overfitting to some extent.
3. *Dropout:* A Regularization method to randomly drop some nodes while training (i.e., setting their weights to 0). This forces the network to learn features in a distributed way. Thus, prevents overfitting and improves generalization.
4. *Flatten:* Flatten layer converts feature maps into a 1D vector so that they can be used for prediction.
5. *Dense layer with Relu:* Dense Layer refers to a simple ANN with a non-linear Relu activation function.
6. *Dense layer with Softmax:* ANN layer with binary activation function Softmax for final classification.

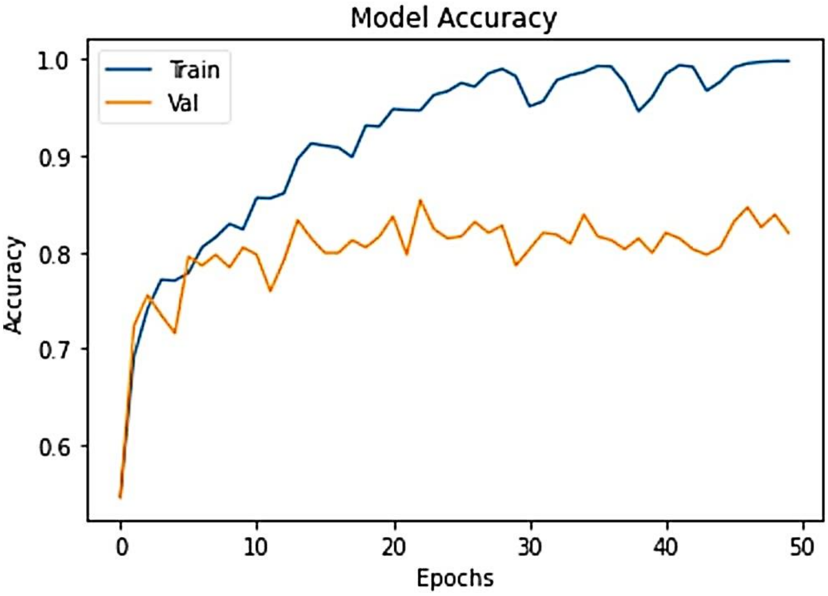

Figure 1. Training model accuracy.

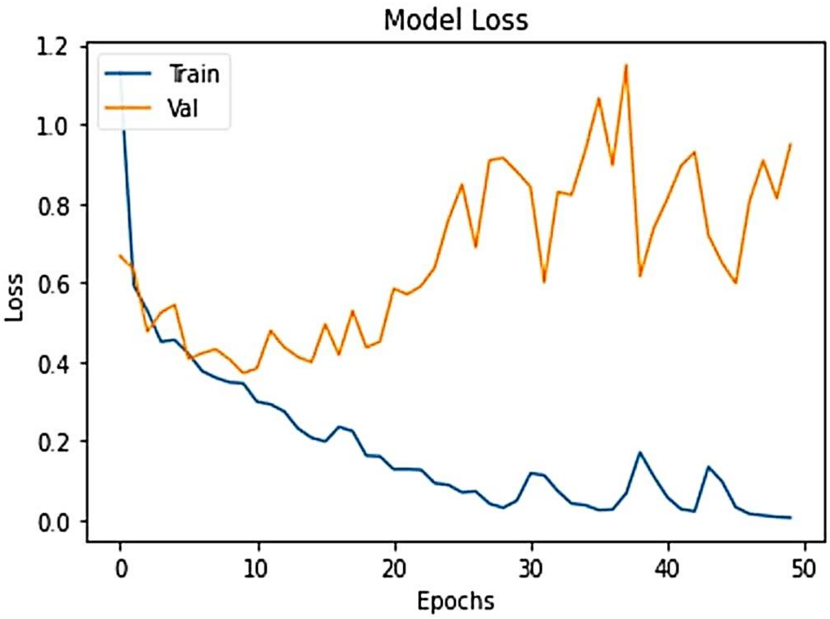

Figure 2. Training model loss.

© STM Journals 2025. All Rights Reserved

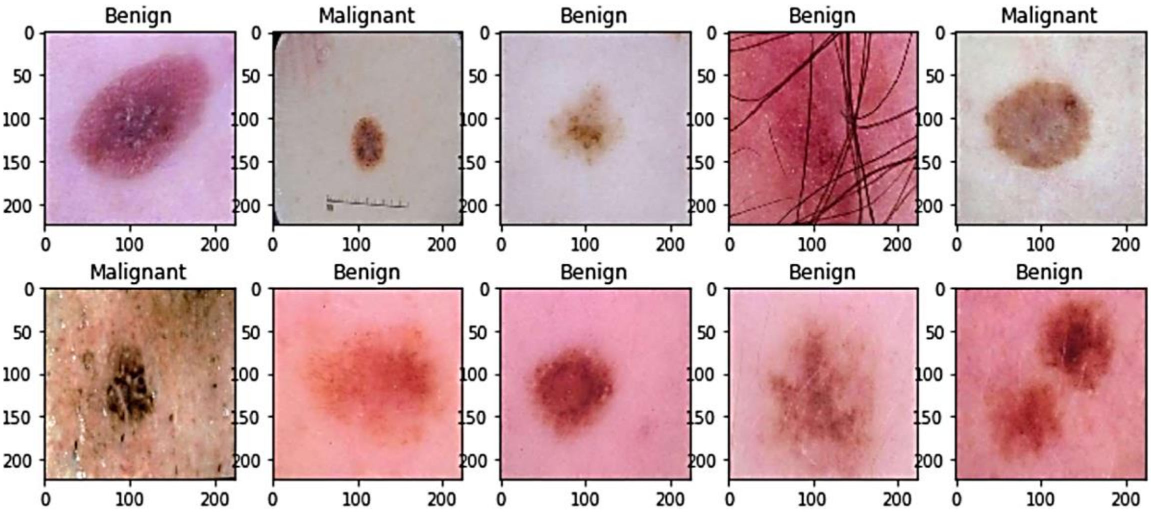

Figure 3. Display benign and malignant images.

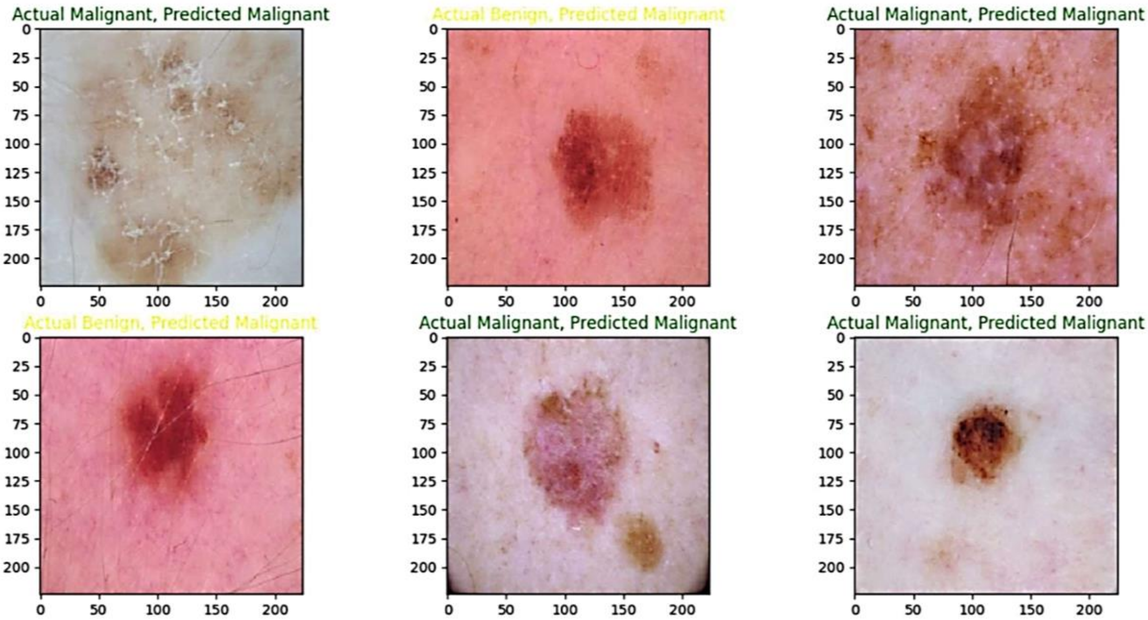

Figure 4. Results are extracted from the basic CNN model.

In our proposed work, the results are visualized using a color-coded scheme to indicate the accuracy and implications of the predictions made by the model. Green images represent correct predictions, where the AI model has accurately identified the condition, whether benign or malignant. Yellow images indicate cases where the model predicted a benign cancer as malignant. While this is technically a misclassification, it is still considered acceptable because it would prompt medical professionals to investigate further, ensuring that no malignant cases are overlooked. Red images, however, highlight the most critical errors, where malignant cancer is incorrectly classified as benign. This is the most concerning scenario, as it could lead to a missed diagnosis, potentially delaying necessary treatment for a serious condition. This color-coding scheme is essential in evaluating the model’s performance and understanding of the potential risks associated with different types of prediction errors.

CONCLUSIONS

This paper addresses the pressing need for early disease detection using sophisticated AIML approaches in medical imaging analysis. This research provided a blueprint based on well-planned data collection combined with preprocessing and exploration analysis, which allowed us to develop robust

© STM Journals 2025. All Rights Reserved

models for determining early characteristics of multiple diseases. Our work utilizes a variety of disease datasets and imaging modalities to reflect the complex nature of disease presentations. We have also relied heavily on clinical experience and domain-specific knowledge to ensure clinical relevance and model interpretability.

The implementation of machine learning techniques, specifically designed for the analysis of medical imaging data, will achieve a major technological shift in disease detection and diagnosis systems. Our approach has been rigorously validated and evaluated to achieve proof of concept and generalization capability, so that it can now be easily translated to medical practice. By enhancing the interpretability of model predictions and providing robust visuals of results, we increase a health professional's ability to make better predictions quickly and, therefore, provide interventions quickly. By enabling early detection and intervention, our approach will have the potential to relieve, lessen health care costs, and improve patient outcomes, reducing the strain on health care systems. In short, this work contributes to the emerging utilization of AI and ML in healthcare and the early detection of disease. As we adapt our methods, we envision a time when AI-enabled systems contribute to the protection of human health and wellness, resulting in a better and healthier tomorrow.

REFERENCES

1. Jaiswal A, Sadana A. Early detection of Alzheimer's disease using bottleneck transformers. Int J Intell Inf Technol. 2022;18(2):1–14.
2. Mahalakshmi K, Sujatha P. The role of exploratory data analysis and pre-processing in the machine learning predictive model for heart disease. In: 2023 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI). Chennai, India: IEEE; 2023. p. 1–8. doi:10.1109/ACCAI58221.2023.10199714.
3. Huang J, Li J, Li Z, Zhu Z, Shen C, Qi G, et al. Detection of diseases using machine learning image recognition technology in artificial intelligence. Comput Intell Neurosci. 2022;2022:5658641. doi:10.1155/2022/5658641.
4. Nasarian E, Sharifrazi D, Mohsenirad S, Tsui K, Alizadehsani R. AI framework for early diagnosis of coronary artery disease: An integration of borderline SMOTE, autoencoders and convolutional neural networks approach. arXiv [Preprint]. Available from: https://arxiv.org/abs/2308.15339.
5. Babu S, Nbs, Kumar V, Divya A, Thanuja B, Ramesh S, et al. AI-driven healthcare: Predictive analytics for disease diagnosis and treatment. Int J Mod Trends Sci Technol. 2024;10:5–9. doi:10.46501/IJMTST1006002.
6. Berkay M, et al. Deep learning based melanoma detection from dermoscopic images. In: 2019 Scientific Meeting on Electrical-Electronics & Biomedical Engineering and Computer Science (EBBT). Istanbul, Turkey: IEEE; 2019. p. 1–4. doi:10.1109/EBBT.2019.8741934.
7. Ghadekar ADP. Early disease detection and prediction using AI technologies: Approaches, future outlook, mitigation strategies, and synthesis of systematic reviews. Int J Intell Syst Appl Eng. 2024;12(3):1434–1445.
8. Haenssle HA, Fink C, Toberer F, Winkler J, Stolz W, Deinlein T, et al. Man against machine reloaded: Performance of a market-approved convolutional neural network in classifying a broad spectrum of skin lesions in comparison with 96 dermatologists working under less artificial conditions. Ann Oncol. 2020 Jan;31(1):137–143.
9. Gautam D, Ahmed M, Meena YK, Ul Haq A. Machine learning-based diagnosis of melanoma using macro images. Int J Numer Method Biomed Eng. 2018;34(5):e2953. doi:10.1002/cnm.2953. PMID:29266819.
10. Fujisawa Y, Otomo Y, Ogata Y, Nakamura Y, Fujita R, Ishitsuka Y, et al. Deep-learning-based, computer-aided classifier developed with a small dataset of clinical images surpasses board-certified dermatologists in skin tumour diagnosis. Br J Dermatol. 2019;180(2):373–381. doi:10.1111/bjd.16924.
11. Kousis I, Perikos I, Hatzilygeroudis I, Virvou M. Deep learning methods for accurate skin cancer recognition and mobile application. Electronics. 2022;11(9):1294. doi:10.3390/electronics 11091294.

© STM Journals 2025. All Rights Reserved

12. Codella NCF, et al. Skin lesion analysis toward melanoma detection: A challenge at the 2017 International Symposium on Biomedical Imaging (ISBI), hosted by the International Skin Imaging Collaboration (ISIC). In: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018). Washington, DC, USA: IEEE; 2018. 68–172. doi:10.1109/ISBI.2018.8363547.
13. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature. 2017;542(7639):115–118. doi:10.1038/nature21056.
14. Yoo TK, Choi JY, Kim HK, Ryu IH, Kim JK. Adopting low-shot deep learning for the detection of conjunctival melanoma using ocular surface images. Comput Methods Programs Biomed. 2021;205:106086. doi:10.1016/j.cmpb.2021.
15. Brinker TJ, Hekler A, Enk AH, Berking C, Haferkamp S, Hauschild A, et al. Deep neural networks are superior to dermatologists in melanoma image classification. Eur J Cancer. 2019;119:11–17. doi:10.1016/j.ejca.2019.05.023. PMID:31401469.
16. Kassem MA, Hosny KM, Fouad MM. Skin lesions classification into eight classes for ISIC 2019 using deep convolutional neural network and transfer learning. IEEE Access. 2020;8:114822–114832. doi:10.1109/ACCESS.2020.3003890.
17. Marchetti MA, Codella NCF, Dusza SW, Gutman DA, Helba B, Kalloo A, et al. Results of the 2016 International Skin Imaging Collaboration International Symposium on Biomedical Imaging challenge: Comparison of the accuracy of computer algorithms to dermatologists for the diagnosis of melanoma from dermoscopic images. J Am Acad Dermatol. 2018;78(2):270–277. doi:10.1016/j.jaad.2017.08.016.
18. Göçeri E. Impact of deep learning and smartphone technologies in dermatology: Automated diagnosis. In: 2020 Tenth International Conference on Image Processing Theory, Tools and Applications (IPTA). Paris, France: IEEE; 2020. 1–6. doi:10.1109/IPTA50016.2020.9286706.
19. Brinker TJ, Hekler A, Enk AH, von Kalle C. Enhanced classifier training to improve precision of a convolutional neural network to identify images of skin lesions. PLoS One. 2019;14(6):e0218713. doi:10.1371/journal.pone.0218713.

© STM Journals 2025. All Rights Reserved