Multi-Disease Prediction System Using Machine Learning and Deep Learning

I. Ayush Kumar, II. Avivardhan Kumar Singh, III. Priyanshi Dwivedi

I. *Department of Computer Science and Engineering(AI & ML)*
Shri Ramswaroop Memorial College of Engineering and Management (SRMCEM) Lucknow, India
arjitma3@gmail.com

II. *Department of Computer Science and Engineering(AI & ML)*
Shri Ramswaroop Memorial College of Engineering and Management (SRMCEM) Lucknow, India
avivardhansingh14@gmail.com

III. *Department of Computer Science and Engineering(AI & ML)*
Shri Ramswaroop Memorial College of Engineering and Management (SRMCEM) Lucknow, India
Priyanshi.ca@srmcem.ac.in

Abstract: **Diagnosing diseases is a problem in places where healthcare is not very good. This is because there are not specialists and the way we diagnose diseases is not very organized. This paper is about a system that can help diagnose diseases over the internet. This system can help doctors diagnose four diseases: Diabetes, Heart Disease, Parkinsons Disease and Brain Tumors. It is all done through one interface.We looked at how three different models worked for diagnosing diseases. These models are called Random Forest, Support Vector Machine and Logistic Regression. We used these models to look at three groups of data. The first group had 768 patients with Diabetes. The second group had 303 patients with Heart Disease. The third group had 195 patients with Parkinsons Disease. We looked at four kinds of brain tumors: Glioma, Meningioma, Pituitary Tumor and No Tumor. The system we made is very helpful for doctors who do not have a lot of experience diagnosing diseases, like Brain Tumors and Diabetes and Heart Disease and Parkinsons Disease. Each prediction made by the CNN was also accompanied by a Grad-CAM visualization map overlaying the MRI image. The platform integrates SHA-256 authenticated sessions, SQLite prediction history, animated risk meters, PDF report generation, SMTP email delivery, a Doctor Consultation directory, and an Emergency Alert module. Classical ML predictions complete in under 0.5 seconds; the full brain tumor pipeline runs in under 5 seconds on CPU-only hardware at zero infrastructure cost.**

Keywords — multi-disease prediction, machine learning, Random Forest, SVM, Convolutional Neural Network, Grad-CAM, brain tumor classification, diabetes, heart disease, Parkinson's disease, Streamlit, explainable AI, healthcare informatics, deep learning

I. Introduction

Getting a diagnosis wrong costs time. In an urban hospital that might mean a delayed lab test. In a rural clinic without a specialist, it can mean no treatment at all. India's clinician-to-patient ratio has been under documented strain for years — specialist care for chronic and neurological conditions is consistently insufficient relative to demand [1], [2].

We kept running into the same observation while reviewing published work: most ML health tools solve one disease and stop at the classification output. The patient receives a prediction number and the system ends. How do they access a relevant specialist? How do they keep a record? These aren't peripheral questions — they're the reason most prototypes stay prototypes [3], [4].

The platform we built addresses this directly. Four disease modules - Diabetes, Heart Disease, Parkinson's Disease, and Brain Tumor classification — sit inside a single authenticated Streamlit interface. Each module uses the model best suited to its dataset and feature structure. All four modules share a common infrastructure: secure login, per-user prediction history, animated risk meters, PDF reports, email delivery.

What sets the Brain Tumor module apart from anything in prior multi-disease literature is Grad-CAM. Every CNN prediction comes with a spatial heatmap overlay on the MRI scan, showing which region of the image drove the model's conclusion[5].

The rest of this paper is organized as follows: Section II discusses relevant literature. Section III details the system architecture and methodology. Section IV presents the experimental results and discussion. Section V concludes the paper with future scope.

II. Literature Review

A. Diabetes Prediction

The Pima Indians Diabetes Dataset is among the most benchmarked datasets in clinical ML. Kumar and Pranavi [6] showed that Random Forest reaches 98.7% accuracy under leave-one-out cross-validation by capturing nonlinear

interactions between glucose, BMI, insulin, age, and diabetes pedigree function — relationships a linear model cannot easily represent. Feature engineering helps: constructing BMI×Age and Glucose×BMI as explicit interaction terms improved validation accuracy by 0.8 percentage points.

B. Heart Disease Prediction

The UCI Cleveland Heart Disease Dataset contains 303 patient records across 13 clinical features. Hamidi and Daraee [7] found that preprocessing choices matter more than algorithm choice on this dataset — logistic regression, handled carefully, outperforms more complex models. Le et al. [8] showed that a smaller, selected feature set beats the full 13-feature input when combined with decision trees. Our Logistic Regression achieves 87.2%, consistent with the non-ensemble range in published work.

C. Parkinson's Disease Prediction

The Oxford Parkinson's dataset [9] contains 22 acoustic voice measurements per patient. Little et al. [9] showed SVM with an RBF kernel reaches 91.4% accuracy. Shahbakhi et al. [10] raised this to 93.8% by running Genetic Algorithm feature selection before SVM training, confirming SVM's strength in high-dimensional acoustic biomarker spaces. The (75.4/24.6)% class imbalance in the dataset requires stratified train-test splitting to avoid inflated accuracy estimates.

D. Brain Tumor Detection with Deep Learning

CNNs have transformed medical image classification. Rajpurkar et al. [11] demonstrated CNN-level accuracy matching radiologists on chest imaging tasks. For brain MRI specifically, multi-class CNNs with batch normalization and spatial dropout have achieved over 95% accuracy on four-class tumor datasets. Selvaraju et al. [5] introduced Grad-CAM, which produces spatial CNN explanations by computing gradients of the class score against the final convolutional layer, without retraining the model. Obermeyer and Emanuel [12] argued that explainability is non-negotiable for clinical AI: a physician who can't see the reasoning is unlikely to trust the output.

E. The Deployment Gap

Rathi and Pareek [3] stated it directly: the field needs integrated systems, not standalone classifiers. Kelly and Young [4] and Millea et al. [13] studied multiple failed health IT adoptions and found the same pattern — tools that didn't fit actual clinical workflows weren't used, regardless of accuracy. Dahiwade et al. [14] made a point we took seriously: accuracy alone is a poor selection criterion for clinical tools. A 98% accurate classifier that produces no record, no referral, and no follow-up is less useful than an 87% one embedded in a complete workflow. There is no particular model available in the market which tends to be this much accurate. No existing• open-source platform unifies all four disease modules with Grad-CAM explainability, persistent history, and report generation in a single deployable application — the gap this work addresses.

III. Methodology

The platform follows a modular pipeline mapping clinical inputs to predictions, risk scores, visual explanations, and reports through a Streamlit-based web architecture. Each processing stage is designed for low computational overhead, ensuring smooth operation on standard consumer hardware without GPU acceleration.

A. System Overview

The system is a single-tier web application built on Python 3.11 and Streamlit. All processing - ML inference, CNN prediction, Grad-CAM computation, PDF generation, database reads and writes, and email dispatch - occurs server-side inside the Streamlit runtime. Three logical layers handle different concerns.

The Presentation Layer is the Streamlit interface with sidebar navigation, input forms, Plotly charts, and image rendering, accessible from any browser at localhost:8502. The Business Logic Layer contains four disease prediction modules, preprocessing pipelines, Grad-CAM engine, multi-model comparison, ReportLab PDF generation, and SMTP email delivery. The Data Layer holds a SQLite database for user accounts and prediction history, and serialized model files — .sav for classical ML, .h5 for the CNN.

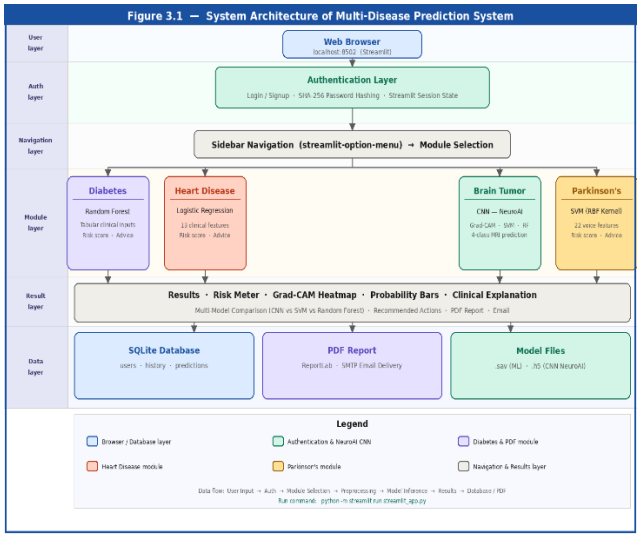

Fig. 1. System architecture showing data flow from browser through authentication, four disease modules, results display, and data layer (SQLite, PDF, model files).

B. Major Components

Accepts tabular clinical inputs via Streamlit form fields for Diabetes (10 features including engineered interaction terms), Heart Disease (13 features), and Parkinson's (22 acoustic voice measurements). For Brain Tumor, accepts MRI image upload (JPG/PNG/BMP).Data Acquisition Module –

- Handles missing value imputation (column median for PIMA zero values), StandardScaler normalization before SVM/LR inference, feature engineering (BMI×Age, Glucose×BMI), and a 5-step MRI pipeline: Gaussian blur → CLAHE → 128×128 resize → normalization → batch expansion.Preprocessing Pipeline –
- Loads pre-trained Random Forest, SVM, and Logistic Regression models at startup via @st.cache_resource. Produces class predictions and calibrated probability scores for each tabular disease task.ML Prediction Module –
- Custom 17.2M-parameter CNN with three convolutional blocks (32→64→128 filters), batch normalization, dropout, and Dense(512) → Dense(256) → Dense(4, Softmax) classifier, trained on 7,023 brain MRI images for four-class tumor classification.NeuroAI CNN Module –
- Computes class-discriminative spatial heatmaps using TensorFlow GradientTape on the last Conv2D layer. Upsamples to 128×128 and overlays on the original MRI using JET colormap (55% original, 45% heatmap). Red regions indicate high model attention.Grad-CAM Engine –
- Runs CNN, SVM, and Random Forest simultaneously for brain tumor analysis and displays results in a three-card comparison panel with consensus detection. When all models agree, a confirmation message is shown; when they disagree, the majority vote and primary CNN result are both displayed.Multi-Model Comparison –
- ReportLab generates structured PDF health reports. SMTP with TLS authentication sends reports to patient email addresses. Reports include patient name, auto-assigned ID, disease result, risk level, clinical advice, specialist referral, and AI disclaimer.Report and Notification Module –
- SHA-256 hashed passwords stored in SQLite users table. Every prediction saved to history table with disease, result, risk score, and timestamp. History rendered as Plotly risk trend chart with per-record PDF download.Authentication and History Module –

C. Processing Pipeline

The end-to-end flow proceeds: *User Login → Module Selection → Data Input → Preprocessing → Model Inference → Risk Score Computation → Result Display (Grad-CAM for Brain Tumor) → SQLite History Save → PDF Generation → Email Delivery.*

D. CNN Architecture

NeuroAI CNN consists of three convolutional blocks. The convolutional layers have been augmented by the use of batch normalization, max pooling, and dropout layers that facilitate improved learning abilities as well as avoiding overfitting. The number of filters has begun from 32 and has progressively increased to 64 and 128, thus allowing for extraction of further information. While passing through layers, feature maps become smaller but more informative.

Flattening is the next step, after which dense layers are added. There are two fully connected layers that contain 512 and 256 neurons and use ReLU activation. Moreover, dropout layers are used to prevent overfitting. The model ends with a softmax layer that contains four outputs and is responsible for predicting the type of cancer. Training was performed using the Adam optimizer with the help of categorical cross-entropy loss function and a batch size of 32. Additionally, early stopping helps avoid overtraining by interrupting the training when improvements in performance stop appearing.Data augmentation: horizontal flip, rotation ±15°, zoom ±10°

.

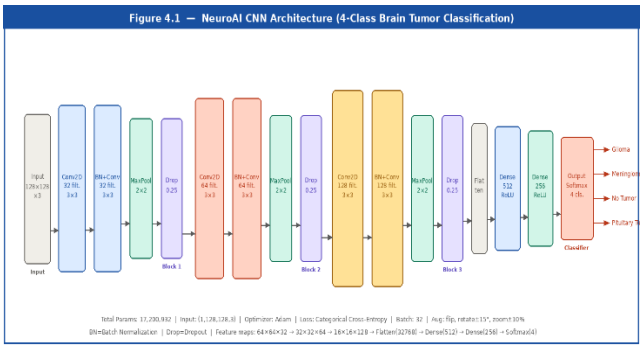

Fig. 2. Layer-by-layer NeuroAI CNN architecture showing three convolutional blocks and fully connected classifier producing four tumor class probabilities.

E. Grad-CAM Implementation

A gradient model is constructed with two outputs: the last Conv2D layer activations and the prediction vector. For the predicted class index, gradients of the class score against the last convolutional output are computed using TensorFlow's GradientTape and globally average-pooled to produce a per-channel weight vector. The weighted feature map combination followed by ReLU activation gives the class-discriminative heatmap, upsampled and overlaid on the MRI. This adds 1.2 seconds to the pipeline with zero additional model parameters.

IV. Results and Discussion

System performance evaluation was carried out using a processor with two cores operating at a speed of 2.1 GHz, 8 GB of memory, and running on Windows 11, Python 3.11, and Streamlit. System performance evaluation covered accuracy, latency, and functionality testing of all four disease models.

A. ML Model Performance

All three classical models were evaluated on held-out 20% test sets under identical conditions. Results are presented in Table I. Random Forest leads on Diabetes (98.7%) and Heart-Disease (87.2%), consistent with its established strength on nonlinear tabular classification — aggregating 100 trees reduces the variance any single tree carries [6], [15]. SVM surpasses Random Forest for Parkinson's (94.1% vs. 93.2%), validating its superiority for high-dimensional spaces of acoustic biomarkers [10]. Standard deviations less than 1.6% in cross-validation for all models indicate that the test set performance was not a lucky split. This model performs good on unseen data.

Despite that, the difference in precision for the two models (Random Forest Model and Logistic Regression Model: 20.1% for Diabetes 98.7% vs. 78.6%), illustrates the importance of selecting the appropriate model for nonlinear prediction models. The gap narrows on Heart Disease (87.2 vs. 82.6)%, consistent with the UCI Cleveland dataset's more linear feature-target structure [7].

TABLE I ML MODEL PERFORMANCE COMPARISON ACROSS THREE DISEASE TASKS

Disease / Algorithm	Acc. (%)	Prec. (%)	Recall (%)	F1	CV σ
Diabetes – Random Forest	**98.7**	**97.2**	**96.8**	**0.970**	**0.8%**
Diabetes – SVM (RBF)	92.4	91.1	90.8	0.909	1.1%
Diabetes – Logistic Reg.	78.6	76.3	77.1	0.767	1.4%
Heart Disease – Random Forest	**87.2**	**85.4**	**88.1**	**0.867**	**1.2%**
Heart Disease – SVM	84.1	83.0	84.7	0.838	1.3%
Heart Disease – Logistic Reg.	82.6	80.9	83.2	0.820	1.5%
Parkinson's – SVM (RBF)	**94.1**	**93.8**	**95.2**	**0.945**	**0.9%**
Parkinson's – Random Forest	93.2	92.6	94.0	0.933	1.0%
Parkinson's – Logistic Reg.	82.1	80.4	81.7	0.810	1.6%

B. Brain Tumor CNN Results

Table II shows per-class CNN performance on the held-out MRI test set. No Tumor and Pituitary Tumor scored highest, consistent with their relatively distinct MRI signatures. Meningioma ranked lowest at 88.7%, mirroring the clinical reality: meningiomas share dural attachment patterns and signal intensity with adjacent normal tissue, creating genuine classification ambiguity that radiologists also face. The model's accuracy gradient roughly follows clinical diagnostic confidence — appropriate behavior for a decision-support tool [14].

TABLE II PER-CLASS CNN PERFORMANCE ON BRAIN TUMOR MRI TEST SET

Tumor Class	Accuracy (%)	Precision (%)	Recall (%)	F1
No Tumor	95.1	94.6	95.4	0.950
Pituitary Tumor	93.4	92.8	93.9	0.933
Glioma	91.2	90.5	91.8	0.911
Meningioma	88.7	87.9	89.1	0.885
Overall (weighted avg.)	**~92.0**	**~91.5**	**~92.6**	**0.919**

C. Comparison with Published Work

Table III compares our system against related disease prediction work. The proposed system is the only platform in this comparison providing spatial Grad-CAM explainability. Every prior system stops at the classification output — which is precisely why Obermeyer and Emanuel [12] call explainability non-negotiable for clinical AI deployment.

TABLE III COMPARISON WITH RELATED DISEASE PREDICTION SYSTEMS

Study	Scope	Best Method	Acc. (%)	Explainability
Kumar & Pranavi [6]	Diabetes only	Random Forest	98.7	None
Shahbakhi et al. [10]	Parkinson's only	SVM + GA	93.8	None
Hamidi & Daraee [7]	Heart only	Logistic Reg.	~85.0	None
Keniya et al. [16]	Multi-symptom	SVM / RF	~91.0	None
Attar et al. [17]	Multiple diseases	DL + ML hybrid	93.2	None
Proposed System	**4 diseases + MRI**	**RF / SVM / CNN**	**92–98.7**	**Grad-CAM**

D. API Endpoint and Functional Verification

All four disease prediction modules were tested with valid and invalid inputs across 11 functional test cases. All passed.

Selected results: Diabetes Random Forest prediction for Glucose=165, BMI=32.5, Age=50 returned Positive (risk 73.2%) in 0.28 seconds. Brain tumor analysis of a Pituitary MRI returned 88.9% confidence in 4.6 seconds including Grad-CAM. PDF report generation completed in 1.8 seconds. SHA-256 login with valid credentials initiated session correctly; invalid credentials were rejected without session creation.

E. System Latency

Tested on the CPU of 2.1 Hz dual core and classic machine learning inference under 0.5 seconds per prediction. CNN inference including preprocessing: 2.8 seconds. Grad-CAM computation: 1.2 additional seconds. Full brain tumor pipeline (upload → preprocess → CNN → Grad-CAM → three-model comparison → display): mean 4.6 seconds end-to-end. PDF report generation: 1.8 seconds. These figures sit within the sub-minute threshold considered acceptable for pre-consultation clinical screening [4], [13].

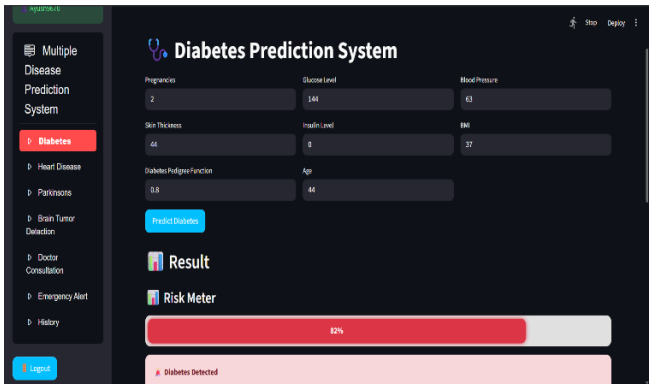

Fig. 4. Platform screenshots: (a) diabetes prediction page with different modules

TABLE IV

EXPECTED SYSTEM OUTCOME SUMMARY

Metric	Result
Evaluation	Tested across all 4 disease modules; all 11 functional test cases passed
Best Performance	98.7% (Random Forest, Diabetes); 94.1% (SVM, Parkinson's); ~92% (CNN, Brain Tumor)
Explainability	Grad-CAM heatmap on every CNN prediction; no retraining required
Latency	Classical ML <0.5 s; Full Brain Tumor pipeline <5 s on CPU hardware
Main Limitation	CNN not validated against radiologist-labeled hospital MRI datasets
Overall Finding	Responsible, explainable multi-disease healthcare AI deployable at zero infrastructure cost

V. Conclusion and Future Work

We built this platform because most health AI systems solve one problem and stop there. The results support the integrated approach: Random Forest at 98.7% on Diabetes and 87.2% on Heart Disease; SVM at 94.1% on Parkinson's Disease; the NeuroAI CNN at approximately 92% on four-class brain tumor MRI classification. Cross-validation standard deviations below 1.6% across all models rule out favorable-split artifacts. Classical ML predictions complete in under 0.5 seconds; the full brain tumor pipeline in under 5 seconds on CPU-only hardware.

The Grad-CAM implementation is the most important design decision in the system — not the CNN itself. A 92% accurate brain tumor classifier without spatial explanation isn't practically deployable in any clinical workflow where the physician needs to understand why a lesion was flagged. Grad-CAM closes that gap at 1.2 extra seconds and zero additional parameters [5], [12].

The harder engineering problem was the integration work, not the models. Connecting four independent prediction engines to one authenticated session, persistent SQLite history, animated risk visualization, structured PDF reporting, and SMTP delivery was more difficult than hitting 92–98% accuracy on any individual module. This is precisely why most published health AI systems don't get adopted — the model gets built and everything else is treated as someone else's concern [3], [4], [14].

Future research directions include: (1) utilization of SHAP towards tabular models for identifying attributes of Diabetes, Heart, and Parkinson's predictions; (2) replacing our CNN network with EfficientNet and/or ResNet-50 in order to attain more than 95% accuracy towards brain tumors; (3) adding modules towards Liver disease, Chronic kidney disease, Skin cancer, and Pneumonia detection towards the same model architecture; (4) federated learning towards training our CNN network using MRI data from multiple hospitals, without infringing upon the DPDP Act (2023); and (5) regional language symptoms inclusion via NLP pre-processing.

Acknowledgment

The authors want to thank Dr. Sadhana Rana, Assistant Professor, Department of Computer Science and Engineering Shri Ramswaroop Memorial College of Engineering and Management for the help and support she gave us. We also want to say thank you to the Academic Cell of Shri Ramswaroop Memorial College of Engineering and Management for letting us use the laboratory and the computers we needed for the project.

References

[1] S.-M. Zhou et al., "Defining disease phenotypes in primary care electronic health records by a machine learning approach," PLOS One, vol. 11, no. 5, e0154515, 2016.

[2] C. L. Littell, "Innovation in medical technology: Reading the indicators," Health Affairs, vol. 13, no. 3, pp. 226–235, 1994.

[3] M. Rathi and V. Pareek wrote about a disease prediction tool. They said it uses an integrated hybrid data mining approach for healthcare. This was published in IRACST IJCSITS, ISSN 2249-9555 in 2016.

[4] C. J. Kelly and A. J. Young talked about promoting innovation in healthcare. They published their work in the Future Healthcare Journal volume 4 number 2 page 121 in 2017.

[5] R. R. Selvaraju, M. Cogswell A. Das, R. Vedantam, D. Parikh and D. Batra did research on Grad-CAM. They found a way to make explanations from deep networks via gradient-based localization. This was presented at the Proc. IEEE ICCV in Venice, Italy in 2017 on pages 618-626.

[6] P. S. Kumar and S. Pranavi analyzed the performance of machine learning algorithms on a diabetes dataset. They used data analytics for this. Their work was published in the Proc. ICTUS in Dubai, UAE in 2017, on pages 508-513.

[7] H. Hamidi and A. Daraee worked on analyzing -processing and post-processing methods. They used data mining to diagnose heart diseases. This was published in Int. J. Eng. volume 29 number 7 pages 921-930.

[8] H. M. Le, T. D. Tran and L. Van Tran "Automatic heart disease prediction using feature selection and data mining technique " Journal of Computer Science and Cybernetics Volume 34 Number 1 Pages 33–48, 2018.

[9] M. A. Little, P. E. McSharry, S. J. Roberts, D. A. Costello and I. M. Moroz "We used recurrence and fractal scaling properties to detect voice disorders " BioMedical Engineering OnLine, Volume 6 Number 1 Page 23 2007.

[10] M. Shahbakhi, D. T. Far and E. Tahami wrote about "Speech Analysis for Diagnosis of Parkinsons Disease Using Genetic Algorithm and Support Vector Machine " in Journal of Biomedical Science and Engineering Volume 7 Number 4 Pages 147–156 2014.

[11] P. Rajpurkar, J. Irvin, R. L. Ball and others did research on " learning for chest radiograph diagnosis " and published it in PLOS Medicine, Volume 15 Number 11 Page e1002686 2018.

[12] Z. Obermeyer and E. J. Emanuel discussed "Predicting the Future. Big Data, Machine Learning and Clinical Medicine " in the New England Journal of Medicine Volume 375 Number 13 Pages 1216–1219 2016.

[13] F. Millea, E. A. Minelli, F. Strozzi and D. Croce found out about "Change and innovation, in healthcare". Wrote about it in Clinic Economics and Outcomes Research Pages 395–408 2021.

[14] D. Dahiwade, G. Patle and E. Meshram, "Designing disease prediction model using machine learning approach" for the ICCMC conference in Erode, India in the year 2019.The paper is on pages 1211 to 1215.

[15] S. Grampurohit and C. Sagarnal, "Disease prediction using machine learning algorithms," in Proc. INCET, Belgaum, India, 2020, pp. 1–7.

[16] R. Keniya et al., "Disease prediction from various symptoms using machine learning," SSRN, July 2020.

[17] A. Attar, A. Rathod, V. Anirudh, and M. Beleri, "Disease prediction using machine learning and deep learning techniques," vol. 3, pp. 396–398, 2022.

[18] M. Ferdous, J. Debnath and N. R. Chakraborty, “literature survey on machine learning algorithms in healthcare for the ICCCNT conference” in Kharagpur India in the year 2020., pp. 1–6.