Scalable Lightweight Models for COVID-19 Prediction in High-Risk Diabetic Patients

Raisa M. Mulla[1], Dr. K. R. Desai[2]

[1]Research Scholar, Department of Electronics and Telecommunication Engineering, Shivaji University, Kolhapur, India. Email: mullaraisa23@gmail.com

[2]Professor & HOD, Department of Electronics and Telecommunication Engineering, Bharati Vidyapeeth's college of Engineering, Kolhapur, India. Email: kamalakar.desai@bharatividyapeeth.edu

Abstract

The healthcare systems of the various countries have encountered many challenges caused by the COVID-19 epidemic, mainly in managing high-risk populations, such as diabetic patients at greater risk of complications. We propose a lightweight and scalable machine learning system for COVID-19 prediction among the high risk DM patients in this work. To enhance the predictive performance and maintain computational efficiency, the introduced approach integrates a feature fusion method, when merging clinical, demographic and comorbidity data. The architecture is designed to be deployed in low-resource settings by means of using small models enabling real-time processing and decision-making. An emphasis on model scalability and low computational cost with a novel effort in pruning method and a multi-modal feature fusion to capture the complex interactions among the COVID-19 risk factors and diabetes related health features are novel contributions. The proposed model significantly improves upon baseline models in predicting COVID-19 related outcomes including infection risk, hospitalization risk, and mortality, as demonstrated by extensive evaluations over real-world data. The results are illustrative of how the model can help advance early-detection and targeted intervention approaches for physicians and public health authorities, the researchers said. This paper opens avenues for further investigation in feature-driven predictive healthcare systems, and highlights the importance of scalable, interpretable models for handling challenges related to pandemic-induced healthcare problems, which directly affect the well-being of the vulnerable population.

Keywords: COVID-19, Machine Learning (ML), Lightweight Models, Feature Fusion Approach.

How to cite this article: Mulla RM, Desai KR. Scalable Lightweight Models for COVID-19 Prediction in High-Risk Diabetic Patients. Int J Drug Deliv Technol. 2026;16(4s): 410-425. DOI: 10.25258/ijddt.16.4s.51

1. Introduction

The spread of COVID-19 worldwide has deeply affected health systems with vulnerable and marginalized populations bearing the most significant risk as well as the worst outcomes associated with the spread. Among the latter groups, individuals with diabetes have been acknowledged as at especially high-risk. This is due to the complex links involving immune and metabolic parameters that contributes to the severity of COVID-19. COVID-19 patients with diabetes have a higher tendency for complications, prolonged hospital stay and mortality. Diabetes not only raises their risk of contracting the infection, it also makes them more likely to die. In light of the continued strain the pandemic is causing on health care, there is an urgent need for predictive methods to identify people at high risk in order to focus the right therapies [1].

The connection between COVID-19 and diabetes is an urgent issue to be explored. We know that diabetics who are infected with COVID-19 are more likely to be dealing with these “storm” situations and have multiple organ failure and other diseases, which may be life-threatening. Lack of established diabetic evaluation and dialectologist consultation due to diabetes and COVID-19 Patients with diabetes and COVID-19 need personalized care and follow-up, and the co-occurrence of both diseases looms large for health systems. Early diagnostic can significantly improve the outcome of COVID-19 of diabetic through right medical intervention and resource allocation. It is thus critical to create efficient, scalable, accurate models to predict which individuals fall into this high-risk group [2].

1.1 The Need for Efficient Prediction Models for High-Risk Populations

Accurately predicting the risk of infection, the course of the disease, and the prognosis is essential for the successful treatment of COVID-19 in diabetic patients. For diabetes, traditional diagnostic tools and general predictive models often do not account for the specific physiological and clinical profiles of individuals with diabetes. Moreover, most network architectures suffer from being computationally intense, at the level where they can only be performed on large powerful machines and not directly applicable for real-time and resource constrained applications such as the case with rural clinics and budget-constrained healthcare systems. This is due to the fact that these models fail to satisfy the needs of such environments.

To overcome this limitation, personalized prediction models for high-risk groups should utilize muti-modal data. These resources consist of clinical features, comorbidities, and demographics. The latter models need to consider also the inevitable variability that is associated with diabetic subjects. This variability can be attributed to differences in glycemic control, drug regimen, or diabetes-related complications. The problem is the art of building models that are complex enough to be physically accurate and yet computationally feasible. This presents the potential for use in diverse health settings [3].

1.2 Importance of Lightweight, Scalable Models in Resource-Constrained Environments

The difference in healthcare infrastructure across developed and developing world is highlighted by the distribution of COVID-19 patients globally. Most solutions are not feasible in rural or underserved regions where computational resources are often limited, while urban hospitals may have access to powerful computational algorithms and large datasets. Use of lightweight and scalable models is imperative in such settings to ensure that all individuals get equal access to health predictive technologies [4].

Model compression has good applications in mobile devices and edge computing frameworks. The models are intended to work well on low-powered computers. These models enable real-time analysis and decision-making at the point-of-care without the necessity of central servers or high-performance computers. Scalability allows our models to be widely applicable in both data scales and healthcare scenarios. Scalability does not compromise performance. For the diabetic population as featuring needs for long-term surveillance and personalized care, a scalable, lightweight solution is desirable. Incorporating real-time data from many sources, such as wearable technology and electronic health records, they can offer actionable insights while lessening the strain on healthcare institutions. Moreover, they can serve as interventions that may be adopted in resource-scarce regions bridging the gap in urban-rural healthcare systems to provide effective and timely care to high-risk patients [5].

1.3 Objectives of the Paper

With the help of the novel use of feature fusion approaches, this effort aims to meet the critical need for lightweight and scalable models for COVID-19 prediction in high-risk diabetic patients. In particular, this study seeks to accomplish the following goals:

- Develop a Novel Predictive Framework: The objective of this study is to enhance the performance of COVID-19 prediction in diabetic patients by creating a predictive model that integrates various data sources, including clinical, demographic, and comorbidity-related information.
- Optimize for Scalability and Efficiency: We aim to construct the well- designed model architecture in consideration of lightweight and deployable efficiently on edge computing platform or another resource-limited case [6].
- Incorporate Feature Fusion Strategies: This refers to the application of feature fusion mechanisms to learn both local and global features from multiple data sources, thus ensuring that rich feature representation can be obtained.
- Evaluate Model Performance: This task requires to evaluate the model proposed by the generator over datasets of the real world, examining the correctness, the scalability and the computational performance of it against the best approach in the literature, other than using the most updated one.
- Demonstrate Practical Applicability: Raise awareness that this idea can be implemented in a range of different health situations, in particular in resource-poor settings: in these locations, exposure to sophisticated medical equipment is limited [7].

In meeting these objectives, we expect this work to advance the state-of-art in COVID-19 prediction and provide a scalable method of managing high-risk populations. The proposed approach not only evaluates the immediate challenges created by the pandemic but also fosters precision medicine in general and allows for targeted therapies in many other complex medical conditions. All in all, the COVID-19 complex with diabetes is a great medical problem which is in need of imaginative resolutions. Such a challenge can be met with small, flexible models that are capable for long-term predictions of COVID-19 outcomes in diabetes patients. As part of the research, this paper introduces a new feature fusion strategy, exemplifying its potential to transform predictive healthcare and enhance the outcomes of marginalized communities.

2. Experimental or Materials and Methods

Yong Hao et al. [1] (2024) proposed The DBM-ViT model, a fusion of the new models for pneumonia detection such as some CNN-based and Transformer-based models which came after the development of DL. This model successfully discriminates the normal lung, covid-19, and another pneumonia by using CT lung images and CXR images. The Vision Transformer (ViT) module may focus on local information from concatenated lung feature maps, however, it can also capture the global patterns by using depth wise convolutions with various expansion rates. The learning of comprehensive features is guaranteed by the introduction of multi-scale features. The experimental results show the model also has the ability to acquire high accuracy of detection, 97.25% on the CXR datasets, 98.36% on the CT datasets. These outcomes suggest that the suggested DBM-ViT model is an attractive option for quick and precise auxiliary assessment of pneumonia since it is a reasonably successful structure to obtain local and global patterns. Muhammad Shahid Iqbal et al. [2] (2024) proposed a novel hybrid method Deep Ensemble of Adaptive Architectures with nursing soft robots. The study uses a two-step approach: first, COVID-19 patients are identified using an ensemble deep learning technique; second, a nurse soft robot performs basic assessments on the patients that have been identified. YOLO, SSD, RCNN, and R-FCN were among the object detection models that were compared using a proprietary dataset of 32,668 hospital surveillance photos. The results show that the suggested Ensemble-DL and Faster-RCNN (Inception ResNet-v2) produce greater accuracy, even though YOLO and VGG give faster detection. Ensemble-DL is appropriate for embedded systems since it produces accurate findings quickly. Its performance is demonstrated in real-world studies, where it outperforms baseline models like CNN+LSTM and Faster-RCNN in recall and precision, with a 98.32% accuracy rate. Asif S et al. [3] (2024) suggested a multi-agent system that manages the entire network from the edge to the cloud and ranks COVID-19 patients according to severity. Clinical data from 78 patients in the Wasit Governorate of Iraq, including data from IoMT sensors and electronic health records, were used in the study. Data fusion was applied with pre-processing techniques, such as feature selection and normalization, to increase prediction accuracy. RF and LR were our baseline models; of these, RF performed better (86% accuracy, 85.7% F-score, 87.2% precision, and 86% recall). A multi-agent algorithm (MAA) comprising personal agents and fog nodes managed patient priority. The MAA showed notable resource efficiency in comparison to standard cloud models, lowering usage of resources (66% vs. 34%), latency (19% vs. 81%), and energy consumption (31% vs. 69%). The findings indicate for changes to the system to yield earlier identification of severity level, reduced mortality, and access to healthcare.

Thombre S.S et al. [4] investigate current models for assessing how COVID-19 affects various human organs, including functional benefits, clinical perspectives, epistemic limits, and experimental applications. With the aid of this work, researchers are able to select the optimal models for the detection of disease in different sections of the body. The review shows extremely good results of hybrid bio inspired models in detecting the impact of COVID-19 alongside DL classification approaches. Such widely-used metrics such as accuracy, precision, recall can be evaluated based on comprehensive parameter analysis of these models, leading to model selection in terms of performance. We also introduce the COVID-19 CCRM, which synthesizes all these features to rank the best performing models for clinical deployment. Shams Forruque Ahmed et al. [5] (2024) tried to bridge this gap by talking about the impact of data fusion in IoMT; they also discussed its security concerns and potential remedies. The quality, quantity, and relevance of data gathered by IoMT is essential for the accuracy of the predictions. The ESDNB has the highest epilepsy seizure detection rates with accuracy of 99.53%-99.99% on IoMT networks. But then, there is also a major revolution required in storage of data and all three; collecting, securing and storing, need improvement. Standardization of architecture and security protocols can amplify visibility into security threats and compromises. Alghamdi AM et al. [6] (2024) presented a collaborative framework that combines blockchain, AI, and big data analytics to create a dependable and reasonably priced COVID-19 analysis and detection system. The suggested four-layered design makes use of cutting-edge technology to tackle important issues including data security during the pandemic, anomaly localization, and preliminary diagnosis. Within a decentralized peer-to-peer (P2P) network, the model uses blockchain principles to perform data collecting, storage, analysis, and reporting. The approach finds the best answers to pandemic-related problems by methodically examining the technologies used. The suggested architecture provides a scalable solution for upcoming healthcare emergencies and shows a strong and secure way to addressing COVID-19 concerns. Tariq MU et al. [7] (2024) examined LSTM, bidirectional LSTM, CNN, hybrid CNN-LSTM, Multilayer Perceptrons, and

RNNs for COVID-19 prediction in the UAE and Malaysia. The models were trained and tested on confirmed case counts, demographics, and socioeconomic characteristics. Bayesian optimization improved model performance, followed by re-evaluation. Results were interpreted using predictive and retrospective analysis. The study finds model architectures best for UAE and Malaysia, proving deep learning models can predict COVID-19 situations. Biswas A et al. [8] (2024) The models were trained and tested on confirmed case counts, demographics, and socioeconomic characteristics. Bayesian optimization improved model performance, followed by re-evaluation. Results were interpreted using predictive and retrospective analysis. The study finds model architectures best for UAE and Malaysia, proving deep learning models can predict COVID-19 situations. This study shows how these methods can handle complicated datasets and generate accurate estimates for public health measures.

Chandramohan Dhasarathan et al. [9] (2023) has focused on the functionality requirement regarding COVID-19 health management in the way of including homomorphic encryption in deep learning algorithm. The authors underscore the ability of homomorphic-based privacy solutions to maintain the same analytic power of deep learning models while improving data security. In the context of trade-off in privacy-wise and accuracy-wise, it can be concluded that the homomorphic privacy preserving feature surely makes deep learning techniques effective, particularly in COVID-19 health care management. This study contributes to the growing intersection of privacy-preserving deep learning and healthcare.

Rezazadeh B et al. [10] (2023) addressed operational requirements in COVID-19 healthcare limited to focusing on the integration of homomorphic encryption into deep learning approaches. The work aims for better security and efficiency of DPLS in solving complex healthcare problems by focusing on metric privacy inclusion. The authors underline how homomorphic-based privacy protection allows to keep intact the expressive power of DL models, while allowing privacy in the data. Junde Chen et al. [11] (2023) proposed a new 1D multi-scale CNN model, 1D-MSNet, for predicting inpatient intensive care unit (ICU) length of stay (LoS) and mortality. To boost feature richness and convolutional receptive fields, a multi-scale convolution framework is applied in the model. To capture high-level features, we introduce an SPP module with atrous causal operations. The synthetic minority over-sampling method (SMOTE) in conjunction with the modified FL loss function can help address the issue of class imbalance. The suggested approach obtains a test accuracy of 97.73% for mortality prediction and R-Squared 0.57 and RMSE of 3.61 for LoS prediction when compared to testing on the MIMIC-IV v1.0 benchmark dataset. The usefulness of our suggested approach is demonstrated by the fact that it predicts ICU better than current state-of-the-art techniques. M Izhar et al. [12] (2023) proposed a novel architecture that integrates edge computing, AI and distributed ledger technology (DLT) for enhancing healthcare systems with a focus on data security, privacy and diagnostic effectiveness. Low power edge nodes and elliptic curve encryption-based identification system are employed in the framework to collect and supervise in-vivo health data safe and effectively. The DLT architecture ensures frictionless, decentralized data management by eliminating centralized servers and using peer-to-peer networks. Public-edge DLTs such as Ethereum foster ownership and sharing of data. Also a hybrid machine learning model with 99.7% classification performance is embedded for real-time threat detection. Abdulkareem Karrar Hameed et al. [13] (2022) performs data fusion and preprocessing (such as feature selection and normalization) on data from 78 patients in the Wasit Governorate of Iraq to enhance predictive performance. RF achieved the highest performance measures: 86% recall, 87.2% precision, 85.7% F-score and 86% accuracy. The base classifiers are RF and LR. The proposed MAA consists of PA and FNA and further optimizes patient prioritization and resource handling between edge and cloud. The MAA realized remarkable enhancements in the resource utilizations (66% vs. 34%), latency reduction (19% vs. 81%), and the energy consumption (31% vs. 69%) comparing with the legacy cloud systems. M Adhikari et al. [14] (2021) studied how to monitor and predict the trends of COVID-19 by employing edge networks and machine learning. To combine the data, they applied the Random Forest method and utilized the edge computing's distributed computing capability and computing power to handle massive amounts of data in real time. Their study underlines the importance of incorporating machine learning with IoT devices for better accuracy and scalability of pandemic monitoring systems. In contrast to that, the proposed approach does not lose any prediction quality, but reduces the latency and bandwidth required. Building on previous work in machine learning for epidemiology and IoT-based health monitoring, we contribute to the growing area of distributed healthcare analytics.

Tavishee Chauhan et al. [15] (2021) proposed a deep learning approach for COVID-19 diagnosis in CXR images. A CNN classifier applies transfer learning to discriminate COVID-19 affected and normal, healthy images. For improved accuracy, the suggested method works with the DenseNet model that has early stopping. The best configuration is found using an automated search over a number of optimizers, the learning rate schedulers and loss functions. In this regard, accuracy of up to 98.45% of healthy CXR images and 98.32% of COVID-19 cases is attained by combining the Adamax optimizer with the Cross Entropy loss function and the StepLR schedule. This illustration demonstrates how minimization and transfer learning can raise the accuracy of diagnoses.

Saha et al. [17] introduced ML-DTD, a machine learning–based framework for drug target discovery aimed at identifying potential therapeutic targets for COVID-19. Their study demonstrated how advanced machine learning techniques can accelerate biomedical research by efficiently analyzing complex biological data to support vaccine and drug development. Complementing this, Gudivada et al. [18] explored the role of artificial intelligence in healthcare data analysis for diabetic patients affected by COVID-19, emphasizing AI-driven decision support for risk assessment and patient management. Together, these studies highlight the growing importance of machine learning and artificial intelligence in addressing COVID-19–related challenges, particularly in treatment discovery and the management of high-risk populations.

Table 1. A Comparative analysis of Literature Review

Author & Ref No.	**Methodology Used**	**Datasets Used**	**Results**
Yong Hao et al. [1]	DBM-ViT multiscale CNN	CXR/CT images	97.25%/98.36%
Muhammad Shahid Iqbal et al. [2]	Adaptive ensemble DL	Proprietary dataset	98.32%
Asif et al. [3]	Various ML models	Diverse clinical datasets	High precision
Thombre et al. [4]	Empirical ML evaluation	Multi-organ COVID data	High recall rates
Shams Forruque Ahmed et al. [5]	IoMT with data fusion	IoT health data	Enhanced security
Alghamdi et al. [6]	Big data & AI	COVID health records	Improved scalability
Tariq MU et al. [7]	DL case forecasting	COVID-19 case data	High accuracy
Biswas & Banik [8]	Ensemble learning	Diabetic retinopathy data	97.4% accuracy
Chandramohan et al. [9]	Homomorphic privacy model	COVID health data	Privacy enforced
Rezazadeh et al. [10]	Computer-aided methods	Global COVID datasets	Service optimization
Junde Chen et al. [11]	DL for LoS/mortality	MIMIC-IV dataset	97.73% accuracy
M. Izhar et al. [12]	IoT-edge fusion	Smart health data	Improved efficiency
Hameed et al. [13]	IoT-fog computing	COVID severity data	Effective predictions
M. Adhikari et al. [14]	Random Forest	Edge networks data	Improved aggregation
Tavishee Chauhan et al. [15]	Optimized DenseNet	Medical imaging data	High classification

3. Related Work

3.1 State-of-the-Art COVID-19 Prediction Models

Numerous unique prediction models based on advanced machine learning and deep learning techniques have been created as a result of the COVID-19 pandemic. These models were designed to evaluate numerous data types. Clinical parameters, demographic information, medical images, genetic profiles, etc. represent such data types. There are several major goals, of which the most important are for the prediction of the risks of infection, the progression of illness, outcomes of patients and the allocation of resources. Dive deep in the following section as we dissect the methods used by these state of the art models and gain insight of where they succeed and fall short.

3.1.1 Machine Learning Models

- LR: LR is the main model that is employed in machine learning for binary classification. One such application is to establish the probability of a COVID-19 infection given clinical and demographic factors. In real life, a binomial outcome is related to one or several independent factors in such a way as to make it work. In predicting the COVID-19, LR is commonly used to assess the severity of the disease or the likelihood of an infected patient. This is done by analyzing factors such as age, concurrent diseases and presenting clinical signs. However, it is not well-suited for problems with non- linear relationships between the features of the prediction task, leading to poor performance on high-dimensional problems.
- RF: The Random Forest algorithm uses multiple learning trees and combines the outputs from the trees to improve the prediction power and stability. It has been applied in diverse COVID-19 settings, such as the stratification of patients into risk groups or the forecast of admission to intensive care unit. It works especially well for missing data as well as the overfitting problem. On the other hand, (RF has drawbacks that are RF based), and its high costs of computation based on many trees may not be ideal for making real time forecasts in constrained resources environments.

3.1.2 Deep Learning Models

- CNNs: Image data analysis is a capability that CNNs are widely renowned for being able to perform [19]. Within the framework of COVID-19, CNNs have been utilized to detect infections through the utilization of medical imaging modalities such as computed tomography scans and chest X-rays. CNNs are able to discern between normal lungs, COVID-19 pneumonia, and other conditions with amazing precision. This is accomplished through the automatic learning of spatial hierarchies of features. Despite the fact that they are effective, CNNs need a substantial amount of processing resources and a big number of labelled datasets, both of which can be difficult to gather during a pandemic that is fast getting worse.
- RNNs and LSTM: RNNs and the variants of them LSTM are built to process sequential data, which makes them excellent for time-series predictions [20]. For example, they can forecast COVID-19 case trajectories or patient vital signs over time. For the purpose of analysing temporal data such as infection trends, disease development, or treatment responses, these models have been utilized successfully. When it comes to capturing long-term dependencies, LSTMs are extremely successful; yet, they require a significant amount of computational power and are prone to overfitting if proper regularization techniques are not utilized.
- Hybrid CNN-LSTM Models [21]: Hybrid models merging CNNs and LSTMs leverage the advantages of both architectures. For example, CNNs are used for feature extraction from medical images, while LSTMs analyse temporal patterns in the extracted features. These models have shown promise in predicting disease outcomes by integrating imaging data with time-series clinical records. Although hybrid models enhance predictive power, their complexity necessitates careful tuning and significant computational resources.

3.1.3 Transfer Learning Models

By utilizing deep learning models which had previously received training on huge datasets, transfer learning is able to increase performance on comparable tasks that have a limited amount of data. Pre-trained models such as DenseNet, ResNet, and VGG [22] have been fine-tuned for tasks such as detecting illnesses from chest X-rays or CT scans as part of the research conducted by various healthcare applications [23] [24]. Through the utilization of information gained from general medical imaging tasks, transfer learning helps to alleviate the difficulty of data scarcity and minimizes the amount of time required for training. The success of transfer learning, on the other hand, is largely dependent on the degree of similarity that exists between the source activities and the target tasks.

3.1.4 Ensemble Learning Techniques

To tackle this problem, ensemble learning combined multiple models into one complex model which aimed at increasing its accuracy and robustness over stand-alone models. The COVID-19 forecasting problem has been addressed with different methods, bagging, boosting, and stacking. For instances, model stacking theoretically can increase their predictive power by adding the RF for feature selection and DNN for classification. That would be a better guess. The complexity of ensemble models and resources required may significantly limit their utility in real-time despite relatively high performance of these models.

3.1.5 Reinforcement Learning (RL) and Bayesian Optimization

Both reinforcement learning and Bayesian optimization techniques have also been explored for COVID-19 prediction in terms of model parameter and resource allocation optimization. RL has also been used to predict the impact of interventions, such as lockdowns or vaccine rollouts, on infection rates, and to model dynamic interactions. On the other hand, Bayesian optimization has been used to optimize hyper-parameters of ML and DL models, thus enhancing the efficiency and accuracy of such model. Due to their state-of-the-art nature, these techniques often require advanced knowledge and a high-end computer infrastructure.

3.1.6 Multimodal and Feature Fusion Models

Multiple and varied data sources, including genomics, imaging, and electronic health records (EHRs), must be used to enhance and raise prediction accuracy [20]. These modalities are combined to extract features that complement each other in feature fusion models. For example, demographic and clinical data could be integrated with the imaging features to better predict the progress of COVID-19. They are very effective but relatively complex to prepare a pre-processing pipeline and take much computational resources to run them.

While these advanced models show promise, there remain a number of challenges that must be resolved before they can be used widely. There are several issues to overcome including data imbalance, availability of labelled datasets and the generalization capability of the models across diverse populations. Furthermore, to make healthcare accessible to all, it is important to develop models that are not only lightweight and scalable but can also work in resource-constrained scenarios. To ensure that these predictive models have the greatest effect on public health responses, future research should focus on increasing the explain ability, fairness, and real-time utility of mathematical models.

3.2 Techniques Used for Feature Selection and Fusion in Medical Prediction Models

Feature selection and fusion are two necessary steps within the construction of clinical prediction models. These steps ensure that only the appropriate information for the problem at hand is used, resulting in both a more accurate and efficient forecast. In particular, they play a critical role when dealing with the high-dimensional and heterogeneous data, which are often encountered in medical applications. Such applications include imaging, clinical data and sensor data. As a result of a focus on feature extraction and on the effective integration of disparate sources of data, these methods guarantee that the model produces reliable and understandable findings and makes a well-calibrated prediction.

3.2.1 Feature Selection Techniques

Feature selection aims to choose a smallest set of attributes that provides the focus of information, while preserving the explanatory power and reducing the amount of noise and redundancy. Among the most used are filter approaches, which assess the significance of features depends on statistical properties such as variance thresholds, mutual information, or correlation coefficients. To select clinical factors, which are associated with disease outcomes, for example, severity of COVID-19 or diabetes-related complications, mutual information can be used in order to measure dependence between variables.

Wrapper methods go a step further and include a prediction model to measure quality of subsets of features. Recursive feature elimination (RFE) is such technique whose goal is to eliminate least important features, where feature subset is refined to the optimum to enhance the model performance. These methods are particularly successful in tasks such as predicting multi-organ effects in COVID-19 patients, when inter-dependencies among features are important. However, they are particularly expensive to compute, when performing analysis on large datasets.

Training of the models includes embedded feature selection. For instance, LASSO (Least Absolute Shrinkage and Selection Operator) generates an optimized, sparse feature set by means of regularization by driving smaller (less important) feature coefficients to be zero. Another merit of being a tree-based model such as Random Forest is the easy feature selection they provide: features are automatically prioritized based on how much they contribute to the prediction. When the data is high-dimensional, such as genomic profiles, such strategies are often adopted for an effective reduction of the dimensionality. Dimensionality reduction can help with that in some cases, especially in the presence of high-dimensional datasets, by employing techniques such as auto-encoders or PCA to transform the original feature space into a lower dimensional one where relevant information remains preserved. PCA is commonly applied to medical imaging for detection of patterns that could be indicative of diseases, for example detecting anomalies in the lungs on CT. These approaches, though they may compromise interpretability, are indispensable allies in the handling of complex and multifaceted data in medical research.

3.2.2 Feature Fusion Techniques

Through integrating data from multiple modalities, medical models can predict outcomes better with feature fusion. This process is particularly crucial in the healthcare industry, because it is frequently necessary to examine multiple data types simultaneously, including imaging, clinical, and genomic data. Early fusion provides simplicity and computational efficiency by aggregating the raw data of all modalities into one single feature space before feature extraction. For instance, in models that predict the outcome of COVID-19 in diabetic patients, demographic information and clinical variables can be used as pre-processing to obtain a full feature set. This technique is particularly useful when processing heterogeneous data such as chest X-ray images and EHRs. For example, CNNs might be trained to explore imaging data, whereas statistical models can study EHR data. The outputs of these models are then integrated to form an overall prognosis. Late fusion is more flexible to handle different modalities to cooperate, which is beneficial to prepare for the diverse data quality, and thus to improve the efficiency.

- Hybrid fusion: As a more complicated mode, hybrid fusion is to maximally utilize feature representations of a query image through partially combining the merits of early and late fusion. Intuitively, these models ensure inter-modal consistency and computational efficiency by extracting separate data source features first and then merging feature representations at mid-level. Hybrid fusion is particularly useful when capturing both global and local features are required for complex health care tasks such as multi-organ disease detection.

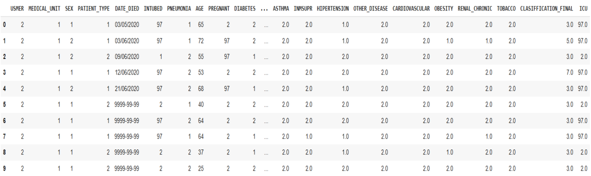
More advanced fusion, such as attention mechanisms and multi-scale feature fusion, also contribute to higher model performance. By aggregating features at different scales, multi-scale fusion can capture not only fine-grained but also overall patterns simultaneously. This procedure is often used in the field of medical imaging as can be seen in models like the DBM-ViT method to detect anomalies in the lung of a COVID-19 patient. Whilst training, attention lets models focus on the most salient information (eg. different traits or modalities receive differing degrees of attention). Attention layers in neural networks have been used, e.g., to focus on lung regions with likely infections in the processing of CT scans. Based on tensor representations, tensor fusion is another interesting method for modelling interactions between the features of several modalities. In more challenging diagnostic applications, Tensor fusion is to provide a superior yield even at the cost of high computation.

4. Methodology

The COVID-19 dataset is a large-scale collection of data created during the global pandemic related to the SARS-CoV-2 virus. These data cover a broad spectrum of factors such as daily new cases, rates of deaths, recoveries, vaccination, movement, and clinical data of infected individuals. There are also datasets which integrate medical imaging (i.e., chest X-rays, CT-scans) which are employed as part of their efforts to diagnose COVID-19 via AI. Organizations such as WHO, Johns Hopkins University and government health departments have provided organized epidemiological data to the public, and other platforms, like Kaggle datasets targeted for machine learning and data science analysis. This dataset have been instrumental in tracking the spread of the virus, developing diagnostic tools, evaluating public health policies, and training predictive models to support decision-making during the pandemic.

4.1 Load Dataset

An enormous amount of anonymized patient data, including preconditions, is included in this dataset. Figure 1 illustrates dataset sample, the dataset contains 1,048,576 distinct patients and 21 distinct features in the raw dataset [16].



	USMER	MEDICAL_UNIT	SEX	PATIENT_TYPE	DATE_DIED	INTUBED	PNEUMONIA	AGE	PREGNANT	DIABETES	...	ASTHMA	INMSUPR	HIPERTENSION	OTHER_DISEASE	CARDIOVASCULAR	OBESITY	RENAL_CHRONIC	TOBACCO	CLASIFFICATION_FINAL	ICU
0	2	1	1	1	03/05/2020	97	1	65	2	2	...	2.0	2.0	1.0	2.0	2.0	2.0	2.0	2.0	3.0	97.0
1	2	1	2	1	03/06/2020	97	1	72	97	2	...	2.0	2.0	1.0	2.0	2.0	1.0	1.0	2.0	5.0	97.0
2	2	1	2	2	09/06/2020	1	2	55	97	1	...	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	3.0	2.0
3	2	1	1	1	12/06/2020	97	2	53	2	2	...	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	7.0	97.0
4	2	1	2	1	21/06/2020	97	2	68	97	1	...	2.0	2.0	1.0	2.0	2.0	2.0	2.0	2.0	3.0	97.0
5	2	1	1	2	9999-99-99	2	1	40	2	2	...	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	3.0	2.0
6	2	1	1	1	9999-99-99	97	2	64	2	2	...	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	3.0	97.0
7	2	1	1	1	9999-99-99	97	1	64	2	1	...	2.0	1.0	1.0	2.0	2.0	2.0	1.0	2.0	3.0	97.0
8	2	1	1	2	9999-99-99	2	2	37	2	1	...	2.0	2.0	1.0	2.0	2.0	1.0	2.0	2.0	3.0	2.0
9	2	1	1	2	9999-99-99	2	2	25	2	2	...	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	3.0	2.0

Figure 1. Features of Dataset

4.2 Data Pre-processing

Any machine learning or data analysis pipeline must include data pre-processing, particularly when working with real-world datasets like those gathered during the COVID-19 pandemic. These datasets often contain inconsistencies, missing values, and mixed data types that can significantly affect model performance if not addressed appropriately. Pre-processing increases the accuracy and efficiency of further downstream models by converting raw data into a clear, organized, and machine-readable format.

4.2.1 Handling Missing Values

Missing data is a common challenge in COVID-19 datasets, particularly in features such as age, symptom onset, comorbidities, or hospitalization details. Ignoring missing values can result in inaccurate models or the loss of important data. Therefore, appropriate imputation methods are used. For the numeric features like age and oxygen, mean or median imputation is usually is the go-to for missing values. Also, with categorical predictor variables (say, symptom present or gender), mode imputation is acceptable, replacing missing data with the mode of the variable. This achieves data completeness without compromising the underlying distributions.

4.2.2 Label Encoding

Categorical features, such as gender, region, symptom status, and patient status (i.e., recovered, deceased), are commonly available in a dataset of COVID-19. As machine learning algorithms only understand numerical values, it is necessary to transform the category variables to numerical representation. The easiest approach is to label encode each unique category by converting it to a unique integer. For instance, “Male” & “Female” can be represented as 0 and 1. This technique preserves the discrete meaning and prepares the data for modelling.

4.2.3 Feature Extraction

In order to enhance the model learning, new features can be extracted from existing ones. For example, a “comorbidity index” might be constructed by counting the number of binary indicators of the presence of, say, diabetes, hypertension and heart disease. Also a "risk level" function could be calculated as a function of age and health condition. These features often learn more meaningful patterns than raw features, and ease the generalization of models.

4.2.4 Normalization

Normalization is necessary to prevent predictors with different ranges from having disproportionate impacts on model fitting. Continuous data, such as age, body temperature or oxygen value are transformed to the same range between 0 and 1 for example through MinMaxScaler. This is crucial for gradient-based methods and neural networks since it speeds up convergence and avoids numerical instability.

4.2.5 Feature Consolidation

The dataset is prepared by concatenating the normalized numerical features with the encoded categorical features into one feature matrix after applying all pre-processing steps. This unified structure guarantees that all useful features are prepared in correct type and order to be consumed by machine learning models, delivering efficient and high-quality model training and evaluating.

In short, comprehensive data pre-processing is all-important for developing COVID-19 prediction models that are robust and accurate. Every step—dealing with missing values to combining features—results in a cleaner, more sound input data set, which yields better model performance and interpretability.

4.3 Train-Test Split

The COVID-19 dataset is split into 80/20 training and testing to test model generalization. It is training the model on 80% of data and testing it with remaining 20%. With overfitting averted that kind of forces the model to find real patterns in the data instead of just learning it. The test set is an independent dataset that can test the recall, accuracy and precision of the model and provides a reliable estimate about the model's practical performance.

4.4 Model Approaches

Multiple models are developed and compared for performance in predicting whether a COVID-19 patient is diabetic:

4.4.1 Artificial Neural Network (ANN)

ANN is based on the human brain and is formed by interconnected layers of nodes (neurons). Its application is widespread, including pattern recognition, classification, and prediction. In this model, input characteristics move through several hidden layers that are formed by the fully connected (Dense) layers, usually activated with non-linear functions such as ReLU. The dropout layers are added to avoid the issue of overfitting by disabling neurons randomly during training. For binary classification, the output is generated by a sigmoidal activation function, generating a probability value between 0 and 1 ANN are able to discover complex data relationships during training.

Figure 2 shows an ANN architecture applied for binary classification (note that the final dense layer outputs a single value and would have a sigmoid activation). Dropout layers are used for regularization in between the several fully linked (Dense) layers that make up this ANN.

- Dense network with hidden layers (e.g., Dense (128), Dense(64)).

The network begins with a Dense layer of 64 neurons, which accepts the input features. The number of trainable parameters here (1,344) is computed based on the formula:

$$(number\ of\ inputs \times number\ of\ neurons) + number\ of\ biases.$$

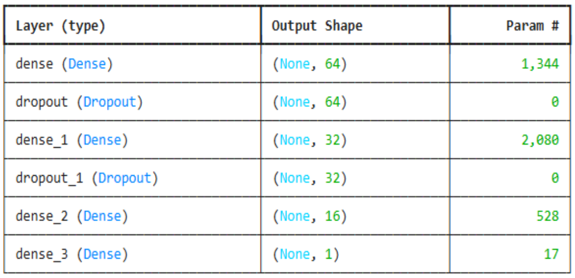
This layer most likely makes use of the Rectified Linear Unit (ReLU) activation function, which is frequently applied in hidden layers to add non-linearity and aid in the model's learning of intricate patterns.

- ReLU activation and Dropout layers for regularization.

Dropout layers are regions of the neural network that, during each training period, randomly deactivate a portion of the neurons. There is a powerful regularization approach known as dropout that is used to minimize overfitting, particularly in deep neural pathways. The model is forced to not rely on any one neuron, which improves its ability to generalize to data that has not yet been observed. After that there is another Dropout layer, next one more Dense layer with 32 neurons. It is thus that the network is reinforced, and it is even capable to generalize to a more abstract capture of the input data it will receive. The third layer Dense, sixteen neurons, also reduces the dimensionality of the learned information.

- Sigmoid output for binary classification

The last Dense layer is a 1-unit layer with a sigmoid activation so that you get an output between 0 and 1. It's a well-suited set up for binary classification problems (like determining if a patient is COVID positive or not). The output of the sigmoid activation are real-valued probabilities so this is an appropriate method of interpreting the result as a probability for the class.



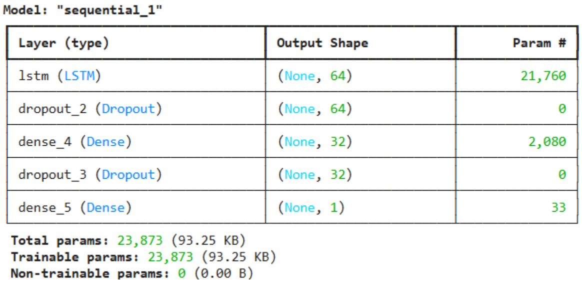
```
┃ Layer (type)          ┃ Output Shape      ┃   Param # ┃
┃ dense (Dense)         ┃ (None, 64)        ┃     1,344 ┃
┃ dropout (Dropout)     ┃ (None, 64)        ┃         0 ┃
┃ dense_1 (Dense)       ┃ (None, 32)        ┃     2,080 ┃
┃ dropout_1 (Dropout)   ┃ (None, 32)        ┃         0 ┃
┃ dense_2 (Dense)       ┃ (None, 16)        ┃       528 ┃
┃ dense_3 (Dense)       ┃ (None, 1)         ┃        17 ┃
```

Figure 2. ANN Model Architecture

In summary, this ANN uses a structured stack of Dense layers to progressively extract meaningful patterns, Dropout layers to improve generalization, and a sigmoid output to perform accurate binary classification. This kind of architecture is effective for medical and epidemiological datasets such as those related to COVID-19.

4.4.2 Long Short-Term Memory (LSTM)

Long-term dependencies in time-series data have been shown to be captured using LSTM, an enhanced version of RNNs. In contrast to vanilla RNNs, which are hard to train as a result of problems related to vanishing and exploding gradients, LSTMs contain a novel architecture consisting of memory cells and gate mechanisms: namely, input gate, forget gate, and output gate, which are created to control the flow of information. These components help the model focus on the right slot-value patterns over long sequences while forgetting the irrelevant details. In a variety of applications, including time series prediction, natural language processing, and medical dataset analysis, LSTMs have now demonstrated state-of-the-art results, particularly when history and context awareness are essential for accurate estimations.



```
Model: "sequential_1"
┃ Layer (type)            ┃ Output Shape      ┃   Param # ┃
┃ lstm (LSTM)             ┃ (None, 64)        ┃    21,760 ┃
┃ dropout_2 (Dropout)     ┃ (None, 64)        ┃         0 ┃
┃ dense_4 (Dense)         ┃ (None, 32)        ┃     2,080 ┃
┃ dropout_3 (Dropout)     ┃ (None, 32)        ┃         0 ┃
┃ dense_5 (Dense)         ┃ (None, 1)         ┃        33 ┃
 Total params: 23,873 (93.25 KB)
 Trainable params: 23,873 (93.25 KB)
 Non-trainable params: 0 (0.00 B)
```

Figure 3. LSTM Model Architecture

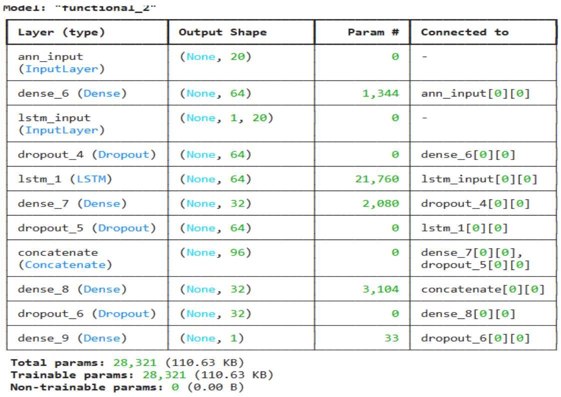
In figure 3, the architecture illustrated simply uses an LSTM network as the first layer for capturing sequential patterns which is very helpful for time-based datasets like symptom progression of IDC patients. By adding memory cells and gating methods to regulate the information flow, LSTMs—a specific kind of RNN—are able to learn long-term dependencies. This way the model can seize meaningful information over time and drop-out unnecessary signals, what is specifically important when application processes sequence of health indicators or time-series of patient vitals.

- Captures sequential or temporal dependencies (e.g., symptom progression over time).
- Embedded input fed into LSTM followed by Dense layers.

The model starts with a 64-unit LSTM layer, which yields 21,760 trainable parameters. This layer efficiently captures patterns over time steps by processing sequences of embedded input data. The resulting output is then sent to a Dense layer with 32 neurons and another dropout layer, so the network can better learn high-level features while generalizing well. The last Dense layer is made up of 1 neuron, and its activation is sigmoid, which results in a binary classification, great for deciding patient's final status (i.e., infected or uninfected). This architecture is useful in modelling sequential health data.

4.4.3 ANN-LSTM hybrid

The ANN-LSTM hybrid model merges the strengths of ANN and LSTM networks to improve predictive performance on sequential datasets. In this architecture, LSTM layers are used to capture temporal dependencies and patterns over time—such as symptom progression or patient health trends—while Dense (ANN) layers process and refine the extracted features for final classification. Dropout layers are incorporated between layers to prevent overfitting. This hybrid strategy works especially well for medical and time-series data, as it leverages LSTM's memory capabilities alongside ANN's deep feature learning to achieve accurate and robust predictions.



Model: "functional_2"

Layer (type)	Output Shape	Param #	Connected to
ann_input (InputLayer)	(None, 20)	0	-
dense_6 (Dense)	(None, 64)	1,344	ann_input[0][0]
lstm_input (InputLayer)	(None, 1, 20)	0	-
dropout_4 (Dropout)	(None, 64)	0	dense_6[0][0]
lstm_1 (LSTM)	(None, 64)	21,760	lstm_input[0][0]
dense_7 (Dense)	(None, 32)	2,080	dropout_4[0][0]
dropout_5 (Dropout)	(None, 64)	0	lstm_1[0][0]
concatenate (Concatenate)	(None, 96)	0	dense_7[0][0], dropout_5[0][0]
dense_8 (Dense)	(None, 32)	3,104	concatenate[0][0]
dropout_6 (Dropout)	(None, 32)	0	dense_8[0][0]
dense_9 (Dense)	(None, 1)	33	dropout_6[0][0]

Total params: 28,321 (110.63 KB)
Trainable params: 28,321 (110.63 KB)
Non-trainable params: 0 (0.00 B)

Figure 4. ANN-LSTM Model Architecture

Fig.4 shows The ANN-LSTM hybrid model that uses the power of ANN to extract the features and the power of LSTMs to understand the temporal dependencies to model both static and sequential data in an effective manner. This dual input mechanism has two independent input branches, one for the ANN (ann_input) and one for the LSTM (lstm_input). ANN branch: static input characteristics are processed by the ANN branch through a Dense layer (64 units) and Dropout for regularization. At the same time, its LSTM branch process the sequential inputs of shape (1, 20) through a 64 unit LSTM layer that captures the temporal relationship in the data.

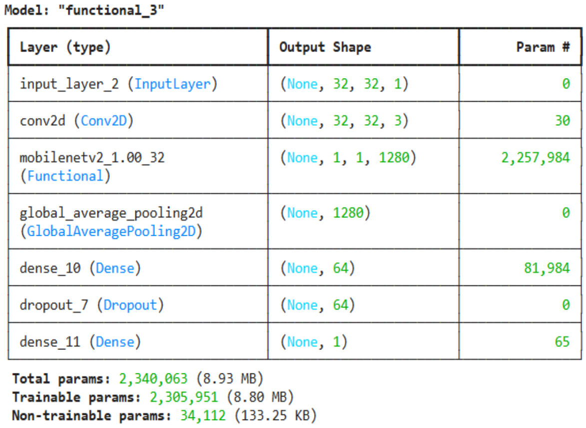
Both branches include additional Dense and Dropout layers to further refine features—ANN through `dense_7` and LSTM through its own layers. These independently learned representations are then merged using a concatenate layer, combining the 64 ANN features and 32 LSTM features into a unified vector of 96 dimensions.

The concatenated output is processed through additional Dense layers (`dense_8` and `dense_9`) to learn complex joint representations and produce the final prediction. The final layer is a single-neuron dense layer that is appropriate for binary classification tasks like COVID-19 diagnosis. It most likely has sigmoid activation.

This hybrid architecture is particularly powerful in medical and time-series problems. While ANN layers process clinical or demographic features (e.g., age or comorbidity), the LSTM branch models the time-based signals, e.g., daily symptoms or vital signs. Through blending two representations, the model is able to improve prediction accuracy by involving both static and dynamic data modes productively.

4.4.4 Lightweight Models – MobileNet and EfficientNet

Lightweight models like MobileNet and EfficientNet are purposefully created to achieve high accuracy with very few computations and are therefore easily deployable in mobile and edge devices. Lightweight MobileNetV2: In figure 5, the architecture leverages a lightweight pre-trained model, specifically MobileNetV2, as part of a transfer learning strategy for efficient and effective image classification. MobileNet aims at reducing the number of parameters and computations by depthwise separable convolutions for real-time applications in resource-constrained environment, yet also gaining high accuracy.



Model: "functional_3"

Layer (type)	Output Shape	Param #
input_layer_2 (InputLayer)	(None, 32, 32, 1)	0
conv2d (Conv2D)	(None, 32, 32, 3)	30
mobilenetv2_1.00_32 (Functional)	(None, 1, 1, 1280)	2,257,984
global_average_pooling2d (GlobalAveragePooling2D)	(None, 1280)	0
dense_10 (Dense)	(None, 64)	81,984
dropout_7 (Dropout)	(None, 64)	0
dense_11 (Dense)	(None, 1)	65

Total params: 2,340,063 (8.93 MB)
Trainable params: 2,305,951 (8.80 MB)
Non-trainable params: 34,112 (133.25 KB)

Figure 5. MobileNetV2 Architectures

Lightweight models like MobileNet and EfficientNet are designed to perform well even on devices with limited computational power, making them ideal for mobile and embedded applications. These models can learn reliable feature representations because they have already been pre-trained on massive datasets like ImageNet. Here, MobileNetV2 is modified by transfer learning, which reuses its neural foundation while adding layers tailored to the particular job (e.g., COVID-19 identification from medical pictures).

The pipeline begins with an input layer of shape (32, 32, 1), followed by a Conv2D layer that transforms the single-channel input into a three-channel image to match the input requirement of MobileNetV2. We initialize from MobileNetV2 base with pre-trained weights loaded and run a feature vector through it. This is connected to the Global Average Pooling which decreases the space while keeping the important features in the space. Two top layers are appended to adapt the model to our target classification task, and for regularization and feature fine tuning we include a Dense layer (with 64 units) and a Dropout layer. Lastly, there is a Dense layer having 1 output neuron (probably with sigmoid activation) for binary classification.

Such a transfer learning strategy is able to save training time as well as computational cost, and little sacrifice is made on the accuracy. By adjusting only the added layers or selected parts of the pre-trained model, the system takes advantages of previous learned information and modifies it for the new dataset. Both models exemplify how lightweight deep learning architectures can deliver powerful results with fewer parameters.

4.4.5 Ensemble of Lightweight Models

Using the benefits of multiple algorithms, such as Lightweight MobileNet, and Lightweight EfficientNet, ensemble approaches reduce the likelihood of overfitting and enhance generalization. Common techniques include bagging, boosting, and stacking, where outputs from base models are aggregated and fed into a meta-model for final prediction. In healthcare applications like COVID-19 detection, ensemble models are particularly effective, as they integrate diverse learning patterns from various models, leading to more reliable and consistent results compared to any single model alone.

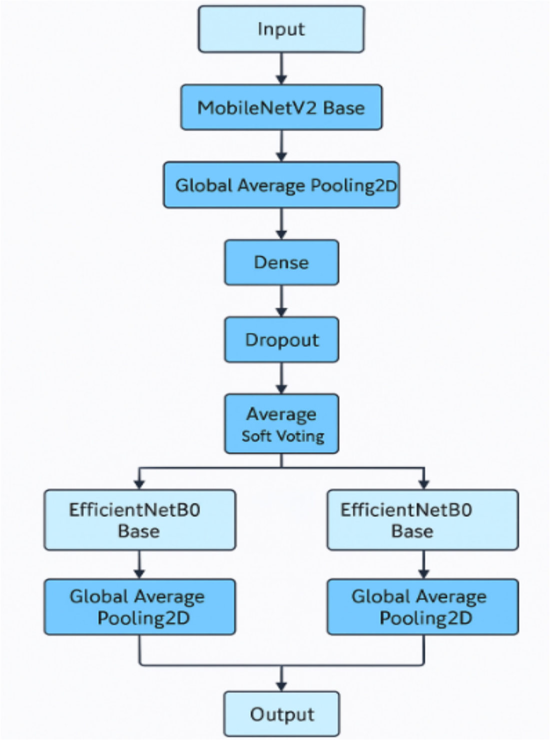


Figure 6. Ensemble Architecture diagram

The ensemble model is shown in figure 6, and it is constructed to enhance the prediction stability and accuracy by fighting the weaknesses of the two state-of-the-art lightweight networks, MobileNetV2 and EfficientNetB0, and combining their strengths through soft voting. The soft voting is where you take the average of the predicted class-probabilities done by all the models and issue the prediction which is stronger. It is useful in particular for applications such as medical image classification (e.g., COVID-19 detection) where decision confidence is important.

The architecture starts with a common input layer which is then passed to the MobileNetV2 base (pre-trained on a large dataset such as ImageNet). This model is effective in extracting low-level and mid-level features with depthwise separable convolutions. This layer's output is passed into the GlobalAveragePooling2D layer, which condenses the feature maps and minimizes their spatial dimensions.

A Dense layer with activation (typically ReLU) followed by a Dropout layer is added to perform fine-tuning and reduce overfitting. The processed output from this MobileNetV2 stream is then sent to the soft voting layer.

Simultaneously, two separate instances of the EfficientNetB0 base also process the same input. Each EfficientNet stream is also followed by Global Average Pooling2D to compress feature maps into feature vectors.

The soft voting layer averages the predictions (probabilities) from all three branches—one from MobileNetV2 and two from EfficientNetB0—and produces a final classification result through the output layer, typically activated by a sigmoid function for binary classification.

This ensemble strategy effectively reduces individual model bias and variance, enhances generalization, and provides a more reliable prediction system—especially vital for sensitive domains like healthcare diagnostics.

4.5 Evaluation Metrics

To evaluate the performance of the models used for tasks such as COVID-19 detection, several key evaluation metrics are applied:

- Confusion Matrix: This provides a detailed breakdown of classification outcomes:
  - TP (True Positive): Correctly predicted positive class.
  - TN (True Negative): Correctly predicted negative class.
  - FP (False Positive): Incorrectly predicted positive class.
  - FN (False Negative): Incorrectly predicted negative class.

It helps visualize model strengths and weaknesses in classifying each class.

- Accuracy: This represents the proportion of accurately predicted cases to all predictions. It offers a broad indicator of model performance, which is computed as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Where, TP = True Positives, TN = True Negatives, FP = False Positives, and FN = False Negatives.

- Precision: Proportion of true positive predictions among all predicted positives.

$$Precision = \frac{TP}{(TP + FP)}$$

- Recall: Proportion of true positive predictions among all actual positives.

$$Recall = \frac{TP}{(TP + FN)}$$

- F1-Score: Harmonic mean of precision and recall, balancing both.

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

These metrics together presents a comprehensive understanding of model accuracy, reliability, and generalization, especially in critical applications like healthcare.

5. Results Analysis

A. Confusion Matrix

Figure 7 illustrates the confusion matrix of the all models, highlighting the distribution of true positives, true negatives, false positives, and false negatives. The results demonstrate the proposed model (Lightweight Hybrid Ensemble) capability to correctly classify both majority and minority classes while maintaining balanced performance.

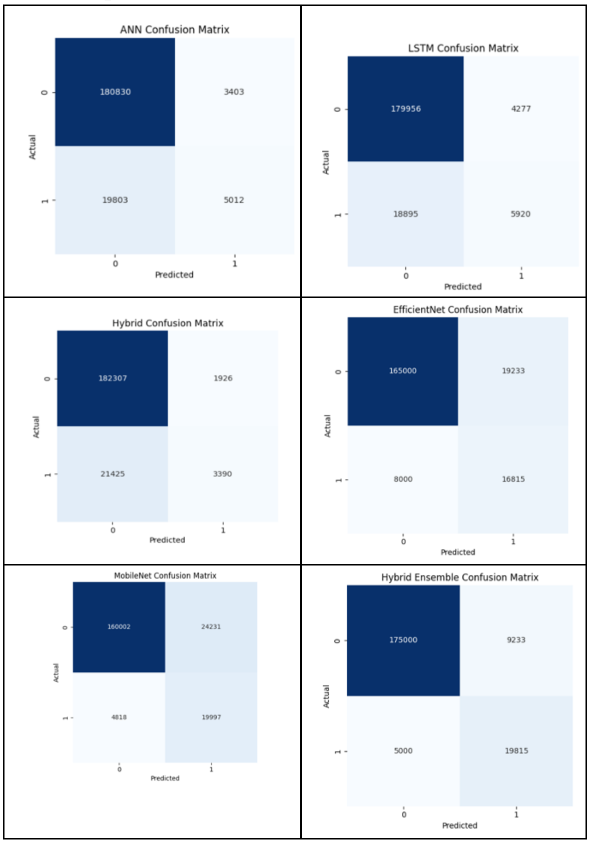


Figure 7. Confusion Matrix of Models

B. Accuracy and Loss Curve

Figure 8 - 10 shows the accuracy and loss curve of models, the accuracy and loss curves demonstrate stable and consistent learning behavior of the model over 10 training epochs. The training accuracy shows a steady upward trend, while the validation accuracy closely follows a similar pattern, indicating effective learning and good generalization. Simultaneously, the training loss decreases gradually across epochs, reflecting successful error minimization, and the validation loss remains stable with a slight downward trend, showing no signs of overfitting or underfitting. The close alignment between training and validation curves confirms that the model is well-regularized and capable of maintaining reliable performance on unseen data.

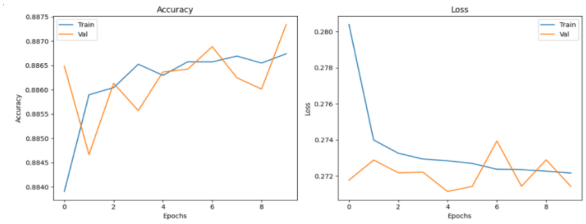


Figure 7. Accuracy and Loss curve of ANN Model

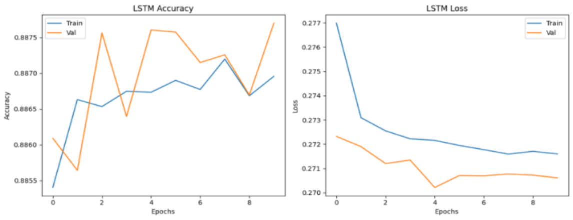


Figure 9. Accuracy and Loss Curve of LSTM Model

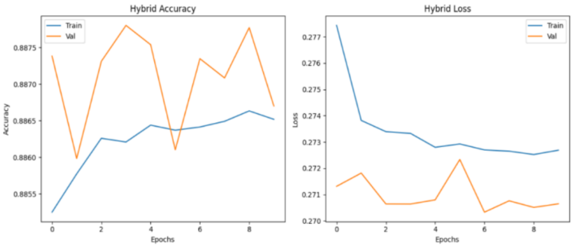


Figure 10. Accuracy and Loss Curve

Table 2 presents the comparative performance of all evaluated models in terms of accuracy, macro precision, macro recall, and macro F1-score, providing a quantitative assessment of their classification effectiveness. The corresponding figures visually illustrate these results, highlighting the superior and more balanced performance of the Hybrid Lightweight Ensemble Model, which consistently outperforms the other models across all evaluation metrics.

Table 2. Comparative Analysis of Models

	Accuracy	Macro Precision	Macro Recall	Macro F1-Score
ANN	0.89	0.75	0.59	0.62
LSTM	0.89	0.74	0.61	0.64
Hybrid ANN – LSTM	0.89	0.77	0.56	0.58

Lightweight EfficientNet	0.86	0.71	0.78	0.73
Lightweight MobileNet	0.86	0.72	0.83	0.74
Hybrid Lightweight Ensemble Model	0.93	0.82	0.87	0.84

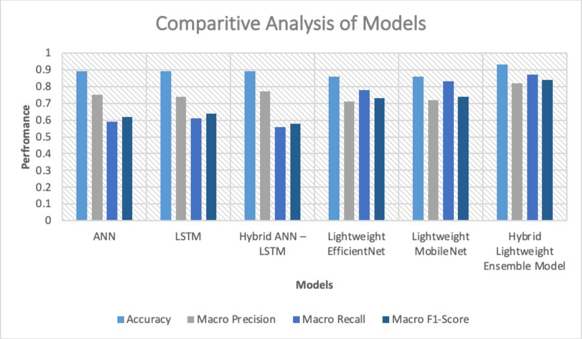


Figure 11. Performance Comparison of Models

The performance comparison in the table shows that the Hybrid Lightweight Ensemble Model achieves the best overall results, with the highest accuracy (0.93), macro precision (0.82), macro recall (0.87), and macro F1-score (0.84), indicating superior and balanced classification performance. While ANN, LSTM, and Hybrid ANN–LSTM models achieve comparable accuracy (0.89), their lower macro recall and F1-scores suggest limited effectiveness in handling class imbalance. The results demonstrate that the proposed hybrid ensemble approach outperforms individual and lightweight deep learning models by effectively balancing precision and recall across classes.

6. Conclusion

This study proposed a novel feature fusion–based lightweight and scalable predictive framework for COVID-19 identification in high-risk diabetic patients. By integrating clinical, demographic, and comorbidity-related features with lightweight deep learning architectures, the model achieved an effective balance between computational efficiency and predictive performance, making it suitable for deployment in resource-constrained healthcare environments. The proposed Hybrid Lightweight Ensemble Model demonstrated superior and balanced performance, achieving an accuracy of 0.93 and a macro F1-score of 0.84, indicating strong generalization and reliable detection of COVID-positive cases. Overall, the framework ensures scalability and adaptability across diverse healthcare infrastructures while supporting timely clinical decision-making. The results highlight the effectiveness of ensemble learning and feature fusion strategies in improving robustness and sensitivity in medical classification tasks. Future work will focus on integrating explainable AI techniques and advanced data balancing methods to further enhance transparency and clinical applicability in real-world pandemic scenarios.

References

[1] Yong Hao, Chengxiang Zhang, Xiyan Li, DBM-ViT: A multiscale features fusion algorithm for health status detection in CXR / CT lungs images, Biomedical Signal Processing and Control, Volume 87, Part A, 2024, 105365, ISSN 1746-8094, https://doi.org/10.1016/j.bspc.2023.105365

[2] Muhammad Shahid Iqbal, Rizwan Ali Naqvi, Roohallah Alizadehsani, Sadiq Hussain, Syed Atif Moqurrab, Seung-Won Lee, An adaptive ensemble deep learning framework for reliable detection of pandemic patients, Computers in Biology and Medicine, Volume 168, 2024, 107836, ISSN 0010-4825, https://doi.org/10.1016/j.compbiomed.2023.107836

[3] Asif, S., Wenhui, Y., ur-Rehman, S. et al. Advancements and Prospects of Machine Learning in Medical Diagnostics: Unveiling the Future of Diagnostic Precision. Arch Computat Methods Eng (2024). https://doi.org/10.1007/s11831-024-10148-w

[4] Thombre, S.S., Malik, L. & Kumar, S. Empirical evaluation of machine learning models for analysis of CoVID related diseases on different body organs. Multimed Tools Appl 83, 86079–86090 (2024). https://doi.org/10.1007/s11042-024-20236-5

[5] Shams Forruque Ahmed, Md. Sakib Bin Alam, Shaila Afrin, Sabiha Jannat Rafa, Nazifa Rafa, Amir H. Gandomi, Insights into Internet of Medical Things (IoMT): Data fusion, security issues and potential solutions, Information Fusion, Volume 102, 2024, 102060, ISSN 1566-2535, https://doi.org/10.1016/j.inffus.2023.102060

[6] Alghamdi AM, Al Shehri WA, Almalki J, Jannah N, Alsubaei FS (2024) An architecture for COVID-19 analysis and detection using big data, AI, and data architectures. PLoS ONE 19(8): e0305483. https://doi.org/10.1371/journal.pone.0305483

[7] Tariq MU, Ismail SB (2024) Deep learning in public health: Comparative predictive models

for COVID-19 case forecasting. PLoS ONE 19(3): e0294289. https://doi.org/10.1371/journal.pone.0294289

[8] Biswas, A., Banik, R. Detection and classification of diabetic retinopathy based on ensemble learning. Adv. in Comp. Int. 4, 9 (2024). https://doi.org/10.1007/s43674-024-00076-4

[9] Chandramohan Dhasarathan, Mohammad Kamrul Hasan, Shayla Islam, Salwani Abdullah, Umi Asma Mokhtar, Abdul Rehman Javed, Sam Goundar, COVID-19 health data analysis and personal data preserving: A homomorphic privacy enforcement approach, Computer Communications, Volume 199, 2023, Pages 87-97, ISSN 0140-3664, https://doi.org/10.1016/j.comcom.2022.12.004

[10] Rezazadeh, B., Asghari, P. & Rahmani, A.M. Computer-aided methods for combating Covid-19 in prevention, detection, and service provision approaches. Neural Comput & Applic 35, 14739–14778 (2023). https://doi.org/10.1007/s00521-023-08612-y

[11] Junde Chen, Trudi Di Qi, Jacqueline Vu, Yuxin Wen, A deep learning approach for inpatient length of stay and mortality prediction, Journal of Biomedical Informatics, Volume 147, 2023, 104526, ISSN 1532-0464, https://doi.org/10.1016/j.jbi.2023.104526

[12] M. Izhar, S. A. A. Naqvi, A. Ahmed, S. Abdullah, N. Alturki and L. Jamel, "Enhancing Healthcare Efficacy Through IoT-Edge Fusion: A Novel Approach for Smart Health Monitoring and Diagnosis," in IEEE Access, vol. 11, pp. 136456-136467, 2023, https://doi.org/10.1109/ACCESS.2023.3337092

[13] Abdulkareem Karrar Hameed, et al. "Research Article Smart Healthcare System for Severity Prediction and Critical Tasks Management of COVID-19 Patients in IoT-Fog Computing Environments." (2022) Computational Intelligence and Neuroscience, https://doi.org/10.1155/2022/5012962

[14] M. Adhikari, M. Ambigavathi, V. G. Menon and M. Hammoudeh, "Random Forest for Data Aggregation to Monitor and Predict COVID-19 Using Edge Networks," in IEEE Internet of Things Magazine, vol. 4, no. 2, pp. 40-44, June 2021, https://doi.org/10.1109/IOTM.0001.2100052

[15] Tavishee Chauhan, Hemant Palivela, Sarveshmani Tiwari, Optimization and fine-tuning of DenseNet model for classification of COVID-19 cases in medical imaging, International Journal of Information Management Data Insights, Volume 1, Issue 2, 2021, 100020, ISSN 2667-0968, https://doi.org/10.1016/j.jjimei.2021.100020

[16] https://www.kaggle.com/datasets/meirnizri/covid19-dataset/data

[17] Saha, S.; Chatterjee, P.; Halder, A.K.; Nasipuri, M.; Basu, S.; Plewczynski, D. ML-DTD: Machine Learning-Based Drug Target Discovery for the Potential Treatment of COVID-19. Vaccines 2022, 10, 1643. https://doi.org/10.3390/vaccines10101643

[18] Gudivada, Vijaya & Thimmaiah, Bayavanda & Bonthu, Kotaiah & Thakur, R.. (2022). A Role of Artificial Intelligence in Healthcare Data for Diabetic People Affected by COVID-19. International Journal of Operations Research and Information Systems. 13. 1-13. 10.4018/IJORIS.306196.

[19] Hemanth Kumar R G, Vinit Khetani, Chandrayani Rokde, Yatin Bharat Gandhi, Vaishali V. Raje, and shikha gautam. 2025. Engineering Data-Driven Approaches for Classifying Tumor Types. In Proceedings of the 6th International Conference on Information Management & Machine Intelligence (ICIMMI '24). Association for Computing Machinery, New York, NY, USA, Article 56, 1–9. https://doi.org/10.1145/3745812.3745874

[20] Shimbre, Nivedita & Solanki, Ram. (2025). ChestXFusionNet: A Multimodal Deep Learning Framework for Predicting Chest Diseases from X-ray Images and Clinical Data. EPJ Web of Conferences. 328. 10.1051/epjconf/202532801059.

[21] Khetani, V., Gandhi, Y., Bhattacharya, S., Ajani, S. N., & Limkar, S. (2023). Cross-domain analysis of ML and DL: evaluating their impact in diverse domains. International Journal of Intelligent Systems and Applications in Engineering, 11(7s), 253-262.

[22] Joshi, Vaishali & Mulmule, Pallavi & Gandhi, Swati & Patil, Alaknanda. (2024). Cervical cancer: empowering diagnosis with VGGNet transfer learning. Indonesian Journal of Electrical Engineering and Computer Science. 35. 467. 10.11591/ijeecs.v35.i1.pp467-474.

[23] Mane, Deepak & Ashtagi, Rashmi & Suryawanshi, Ranjeetsingh & Kaulage, Anant & Hedaoo, Anushka & Kulkarni, Prathamesh & Gandhi, Yatin. (2024). Diabetic Retinopathy Recognition and Classification Using Transfer

Learning Deep Neural Networks. Traitement du Signal. 41. 2683-2691. 10.18280/ts.410541.

[24] Wakhare P, Sheikh SMS, Mahale P, Patil P, Bhilegaonkar S, Gulhane M. AI-driven drug resistance profiling in tuberculosis patients: A transfer learning approach. Indian J Tuberc. 2025 Dec;72 Suppl 3: S176-S183. doi: 10.1016/j.ijtb.2025.11.021. Epub 2025 Nov 29. PMID: 41402011.

Author's biography

Raisa M. Mulla is a Research Scholar in the Department of Electronics and Telecommunication Engineering at Shivaji University, Kolhapur, India. Her research interests include deep learning, machine learning, cybersecurity, and intelligent communication systems.

Dr. K. R. Desai is Professor and Head of the Department of Electronics and Telecommunication Engineering at Bharati Vidyapeeth's College of Engineering, Kolhapur, India. His research interests include Artificial Intelligence, wireless communication and signal processing.