International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

Deep Genomic Signature Classifier for Covid-19 Strains

I. Shalini[1*], D. Apeksha[1], Shaik Anuf[2], Shaik Hasheera[2], Yakasiri Lavanya[2], Thalamanchi Lasya Priya[2]
[1]Assistant Professor, [2]UG Student, [1,2]Department of Computer Science and Engineering
[1,2]Geethanjali Institute of Science and Technology, Nellore-Bombay Highway, S.P.S.R, Andhra Pradesh 524137, India
*Correspondence: I. Shalini (shalini@gist.edu.in)

ABSTRACT
The rapid growth of biomedical data and infectious disease monitoring has created a strong demand for intelligent systems capable of accurately predicting viral status based on patient and sequencing-related attributes. Early and precise detection of viral infections plays a crucial role in disease control, treatment planning, and public health management. However, traditional diagnostic approaches and conventional data analysis techniques often struggle to efficiently handle high-dimensional and heterogeneous datasets. Existing traditional systems primarily rely on statistical methods or single machine learning models for classification. While these approaches provide baseline performance, they often lack robustness, fail to capture complex non-linear relationships in data, and are sensitive to feature variations. Additionally, standalone models may not generalize well across diverse datasets, leading to reduced prediction accuracy and reliability. These limitations highlight the need for a more advanced, flexible, and data-driven approach. To address these challenges, the proposed system introduces a multi-model intelligent framework that integrates several machine learning (ML) algorithms such as Random Forest (RF), Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Logistic Regression (LR), Decision Tree Classifier (DTC), Gradient Boosting Classifier (GBC), and Adaptive Boosting Classifier (ABC), alongside a Proposed Hybrid Deep Learning (PHDL) model based on a Dense Batch Normalization Serial Neural Network (DBN-SNN). In addition, an Optimal Rule List Classifier (ORLC), implemented as a RF based interpretable model, is incorporated to enhance decision-making. The system adopts a performance-based selection strategy, where both the PHDL model and ORLC are trained and evaluated, and the best-performing model is selected dynamically for final prediction. Furthermore, the system is deployed as a web-based application using Flask, enabling data upload, preprocessing, exploratory data analysis, model training, performance comparison, and batch prediction. This framework improves accuracy, scalability, and supports data-driven healthcare analytics.

Keywords: Machine Learning (ML), Deep Learning (DL), Ensemble Learning, Healthcare Analytics, Infectious Disease Monitoring

1.INTRODUCTION

The first cases of COVID-19 were reported in December 2019 in Wuhan, China, marking the beginning of what would soon become a global pandemic. Since then, the disease has spread rapidly across countries and continents, resulting in an unprecedented public health crisis. As of January 2023, more than 668 million confirmed cases and over 6.7 million deaths have been reported worldwide, highlighting the severe global impact of the SARS-CoV-2 virus [1]. The clinical manifestations of SARS-CoV-2 infection vary widely, ranging from asymptomatic cases to severe and life-threatening conditions. Common symptoms include fever, cough, fatigue, diarrhea, shortness of breath, pneumonia, and acute respiratory distress syndrome (ARDS) [2]. In severe cases, patients may require intensive care support, including mechanical ventilation. The severity of COVID-19 is strongly associated with underlying health conditions. Individuals with comorbidities such as diabetes, cardiovascular diseases, hypertension, and chronic respiratory disorders are at a significantly higher risk of severe illness and

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

mortality. These conditions are particularly prevalent among the elderly population, making age a critical risk factor [3]. Beyond age and comorbidities, several studies have identified additional determinants influencing disease progression, including viral genetic variations, host genetic susceptibility, and biological sex differences, which may affect immune response and disease outcomes [4].

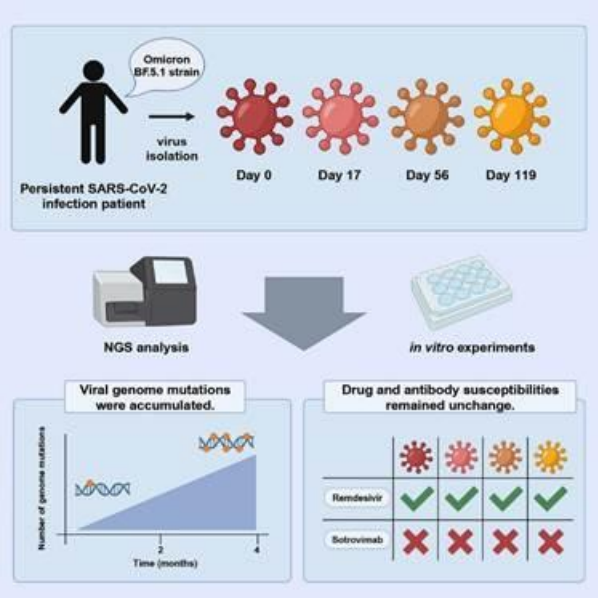


Fig. 1: Longitudinal analysis of genomic mutations in SARS-CoV-2 isolates from persistent COVID-19 patients.

The pandemic has not only caused widespread morbidity and mortality but has also led to substantial economic disruption, healthcare system overload, and social challenges globally. Although the development and large-scale deployment of effective vaccines—such as the mRNA-based Pfizer-BioNTech and Moderna vaccines—have significantly reduced infection rates and severity in many regions, the virus continues to evolve. Emerging variants with mutations in their genomic structure pose ongoing challenges to vaccine efficacy and disease management. Therefore, continuous analysis of COVID-19 genomic sequences remains critically important. Genomic surveillance enables the identification of new variants, understanding of transmission patterns, and assessment of mutation impacts on virulence and immune escape. This information is essential for improving predictive models, guiding public health strategies, and identifying high-risk individuals, ultimately supporting more effective prevention and control measures [5].

2 LITERATURE SURVEY

A. M. Mutawa et, al. [6] introduced a deep learning-based COVID-19 genomic sequence categorization approach. Attention-based hybrid deep learning (DL) models categorize 1423 COVID-19 and 11,388 other viral genome sequences. An unknown dataset is also used to assess the models. The five models' accuracy, f1-score, area under the curve (AUC), precision, Matthews correlation coefficient (MCC), and recall are evaluated Jemila Deida et, al. [7] The purpose of the present study was to document the genomic pattern of SARS−CoV−2 variants from clinical isolates during the COVID−19 outbreak in Mauritania, from September to November 2021. The whole genomes from 54 SARS−CoV−2 strains detected in nasopharyngeal swabs with a cycle threshold value ≤ 30 were successfully sequenced using next−generation sequencing (NGS) and the

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

Illumina protocol. The mean genome coverage (±standard deviation) was 96.8% (±3.7). The most commonly identified clade was 21J (57.4%), followed by 21D (16.7%), 20A (11.1%), and 20B (9.2%). Karolayne S Azevedo et, al. [8] presented Deep Virus Classifier, a tool capable of accurately classifying Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) viral sequences among other subtypes of the coronaviridae family. This classification is achieved through a deep neural network model that relies on convolutional neural networks (CNNs). Kuganya Nirmalarajah et, al. [9] A cohort of adult patients with laboratory confirmed SARS-CoV-2 from 11 participating healthcare institutions in the Greater Toronto Area (GTA) were recruited from March 2020 to April 2022. Supervised machine learning (ML) models were developed to predict hospitalization using SARS-CoV-2 lineage-specific genomic signatures, patient demographics, symptoms, and pre-existing comorbidities. The relative importance of these features was then evaluated. Aruna Rajalingam et, al. [10] Their study reported the transcriptomic profile of the long COVID patient's whole blood samples that are collected from 0 to 35th day of acute infection as described in the GSE215865 dataset (1391 Samples after preprocessing: 1233-COVID positive and 158-COVID negative). The binary classification algorithm from the sci-kit learn python library, namely logistic regression and random forest with 10-fold cross-validation, was applied to the processed data, followed by a selection of the 20 best gene features with recursive feature elimination from a set of 10,719 gene features to obtain the classification accuracy of 87%.

Rajkumar Pandiarajan et, al. [11] presented a comprehensive machine learning framework, starting with advanced preprocessing techniques, including K-mers, one-hot encoding, and Principal Component Analysis (PCA), to ensure high-quality input data and effective dimensionality reduction. For feature extraction, the framework integrates a Dual-Crossed Squeeze Net and a Self-Adaptive Harris Hawk Optimization (SA-HHO) algorithm with a lookahead optimizer to enhance feature learning and representation. The classification stage employs Support Vector Machine (SVM) and Random Forest models, combined using a majority voting mechanism to achieve robust and accurate predictions Shivendra Dubey et, al. [12] presented an approach that used SARS-CoV-2 mNGS (meta-genomic next-generation sequencing) samples to apply XAI (explainable artificial intelligence) methodologies derived from the use of machine learning methods. The research's data set contained 15,979 gene expression profiles from 234 infected individuals, of whom 39.68% (93) tested positive for SARS-CoV-2 and 60.29% (141) tested negative. The SARS-CoV-2–related genes were selected using the LASSO (least absolute shrinkage and selection operator) technique. The class imbalance issue was resolved using the SVM-SMOTE (Support Vector Machine – Synthetic Minority Oversampling Technique) approach. To identify potential SARS-CoV-2–related biomarkers and enhance the interpretability of the final model, an explainable strategy that utilized SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-Agnostic Explanations) methods was employed. The XGBoost model (achieved 92.98% accuracy) and performed better in identifying infectious diseases like COVID-19 compared to the LR (achieved 91.17% accuracy), SVM (achieved 87.68% accuracy), and RF (achieved 91.18% accuracy) models. The three most significant genes associated with COVID-19 according to the SHAP method were FAM83A, LGR6, and IFI27.

3. PRPOPOSED METHODOLOGY

The system begins by loading structured data from a CSV file and organizing it for processing. The data is preprocessed using label encoding for categorical features and standardization for numerical values, followed by splitting into training and testing sets. Exploratory Data Analysis is performed to understand data distribution, relationships, and potential patterns. Multiple machine learning models,

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

including Random Forest, SVM, KNN, LR, DTC, GBC, and ABC, are trained for comparison. A proposed hybrid approach is implemented using a DBN-SNN model for deep feature learning and an ORLC model based on RF for classification. Both models are trained independently, and the one with higher test accuracy is selected as the final deployed model. All models are evaluated using metrics such as accuracy, precision, recall, and F1-score, and the results are stored for analysis. The system is deployed using a Flask web application that provides modules for EDA, classification, performance comparison, and prediction. Users can upload new CSV data through the interface, where it undergoes the same preprocessing using saved encoders and scaler. Predictions are generated using trained models, and results are appended to the dataset and made available for download as illustrated in Fig. 2.

Step 1: Data Input: The system starts by loading the dataset containing structured input features. The data is uploaded and organized in a tabular format for processing. This step ensures that raw input is available for further analysis.

Step 2: Data Preprocessing: The data is transformed using label encoding and feature standardization techniques. It is then split into training and testing sets for model development. This ensures consistency and prepares data for accurate learning.

Step 3: Machine Learning Models: Multiple models such as RF, SVM, KNN, LR, DTC, GBC, and ABC are trained. Each model learns patterns from the training dataset. These models are used for comparison and evaluation.

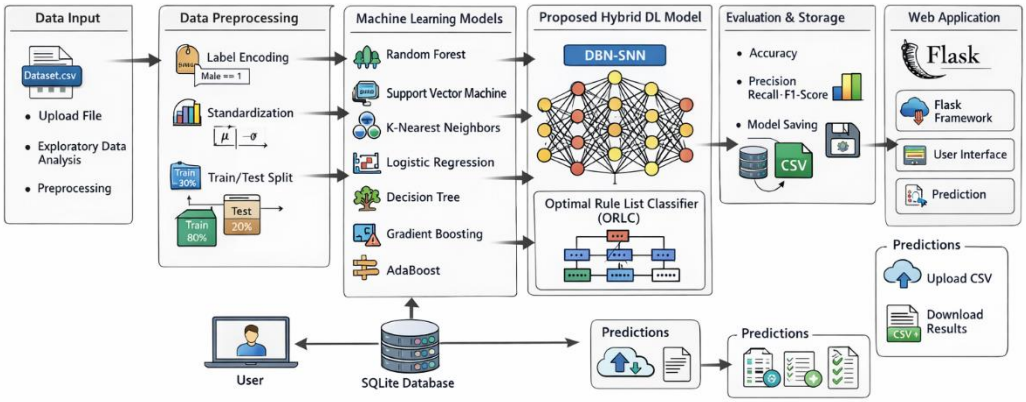


Fig. 2: Proposed System Architecture

Step 4: Proposed Hybrid DL Model: The DBN-SNN model extracts deep features using multiple dense layers. ORLC performs classification using a Random Forest-based approach with rule summaries. The model with better accuracy is selected as the final output model.

Step 5: Evaluation & Storage: All models are evaluated using accuracy, precision, recall, and F1-score metrics. Confusion matrices and classification reports are generated for analysis. The trained models and results are stored for future use.

Step 6: Web Application (Flask): The system is deployed using a Flask web framework for user interaction. It provides modules for EDA, classification, performance, and prediction. This enables easy access to all system functionalities.

Step 7: Prediction Pipeline: Users upload new CSV data through the web interface for prediction. The system applies the same preprocessing using saved encoders and scaler. Predictions are generated using all trained models including the hybrid model.

Step 8: Output & Results: The predicted results are appended to the input dataset for user reference. The final output is displayed on the interface and available for download. This ensures easy interpretation and usability of predictions.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

4. RESULTS AND DISCUSSION

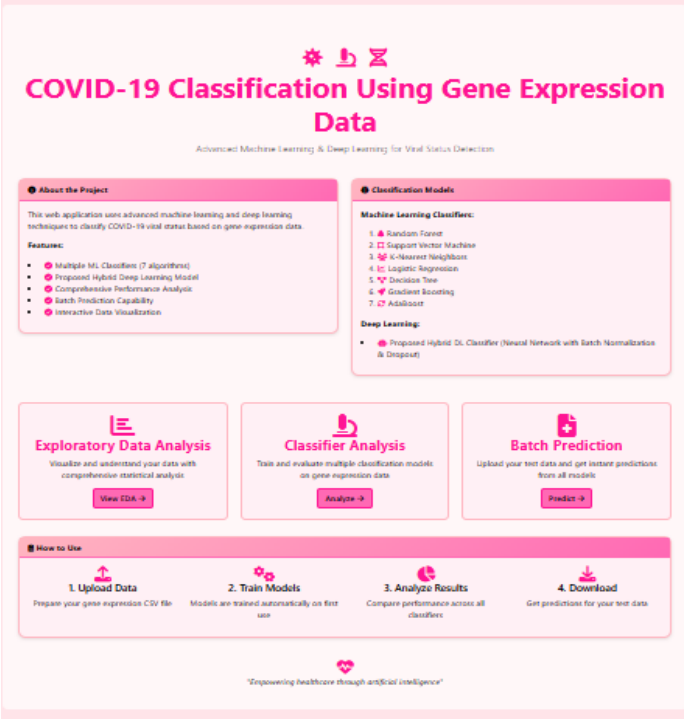


Fig. 3: User Interface of COVID – 19 strains classification.

Fig. 3 illustrates the main user interface of the COVID-19 classification web application, designed with a clean and intuitive layout. It presents key sections such as Exploratory Data Analysis, Classifier Analysis, and Batch Prediction, each accessible via prominent buttons. The interface highlights the use of advanced machine learning and deep learning techniques for viral status detection from gene expression data. It also lists available models including like RF, SVM, and the PHDLC. Interactive data visualization and model performance comparison features are emphasized for user-friendly analysis.

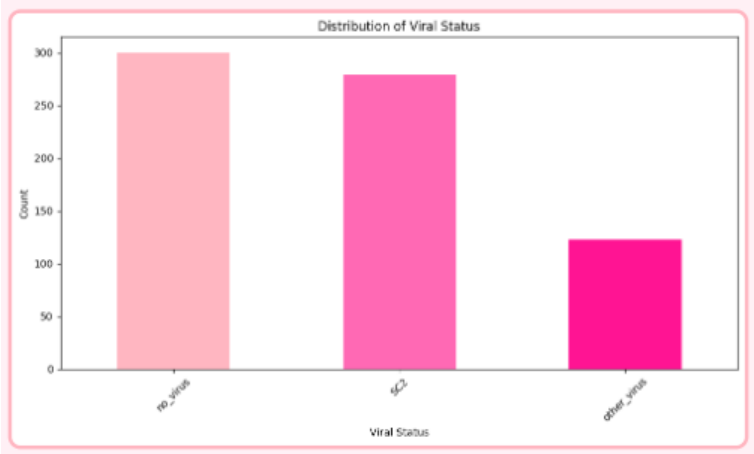


Fig. 4: Viral status distribution.

Fig. 4 presents a bar chart depicting the distribution of viral status across the dataset. It shows three distinct categories: "no_virus," "SC2," and "other_virus," with "no_virus" and "SC2" having nearly equal high counts around 290–300 samples each. The "other_virus" category is significantly smaller, with approximately 120 samples. The chart uses a consistent pink color scheme and clear labeling on

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

both axes. This visualization effectively communicates class imbalance and the dominance of SARS-CoV-2 (SC2) and negative cases in the dataset.

Fig. 5 shows a pie chart illustrating the gender distribution within the study cohort. It reveals that 53.0% of the participants are female (F), represented in a lighter pink shade, while 47.0% are male (M), shown in a darker pink tone. The chart is clearly labeled with percentage values directly on the segments for immediate interpretation. The near-equal split suggests a balanced representation of both genders in the dataset. This visualization supports demographic analysis and ensures transparency in sample composition for clinical interpretation.

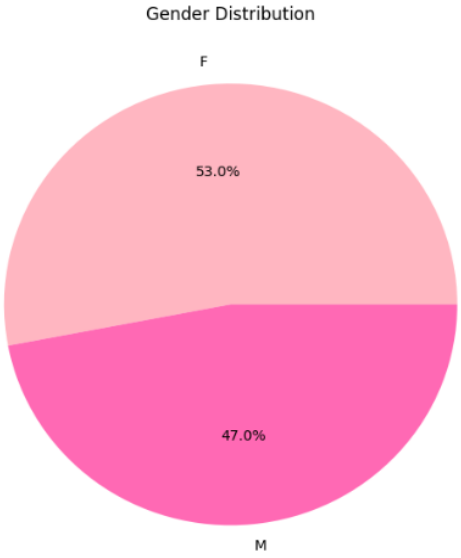


Fig. 5: Pie Chart visualizing Gender Distribution.

Fig. 6 illustrates the Proposed Hybrid DL Classifier's perfect performance, with a diagonal confusion matrix: 56 for Actual 0, 60 for Actual 1, and 25 for Actual 2. The report confirms 1.00 precision, recall, and F1-scores across "no_SC2" and "other_virus," with full 1.00 weighted accuracy. This visualization validates the hybrid deep learning approach's superior classification capability.

Proposed Hybrid DL - Detailed Metrics

Confusion Matrix (CM):

	Predicted 0	Predicted 1	Predicted 2
Actual 0	56	0	0
Actual 1	0	60	0
Actual 2	0	0	25

Classification Report (CR):

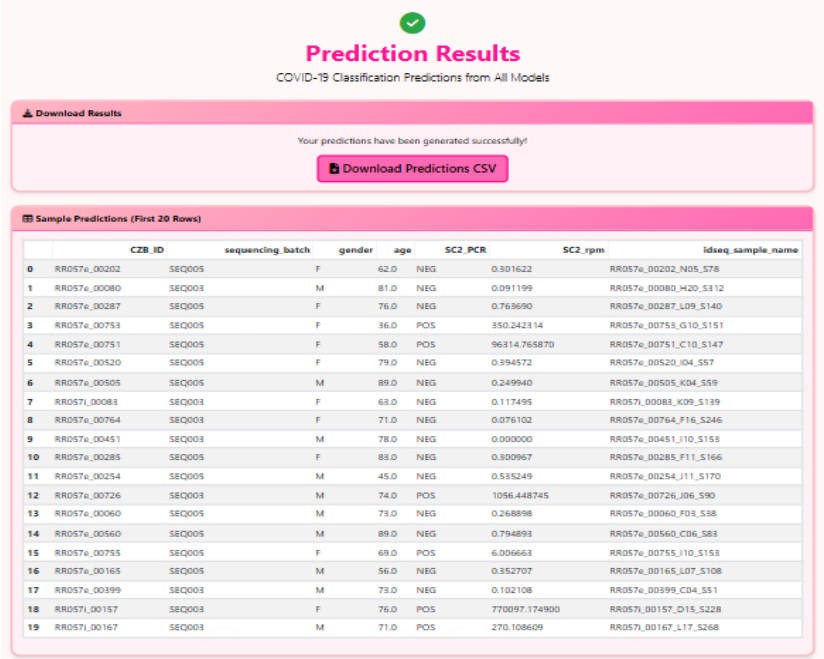
	precision	recall	f1-score	support
SC2	1.00	1.00	1.00	56
no_virus	1.00	1.00	1.00	60
other_virus	1.00	1.00	1.00	25
accuracy			1.00	141
macro avg	1.00	1.00	1.00	141
weighted avg	1.00	1.00	1.00	141

Fig. 6: Confusion matrix for Proposed Hybrid DL Classifier.

Fig. 7 illustrates the Prediction Results interface, displaying a successful batch prediction outcome for COVID-19 classification across 234 test samples. It presents a preview table of the first 20 rows, including key clinical features such as CZB ID, sequencing batch, gender, age, SC2 PCR status, SC2 rpm, and sample name. The interface confirms generation completion with a green checkmark and

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

offers a prominent "Download Predictions CSV" button for full result export. This section ensures transparency by showing raw input data alongside predicted outcomes, facilitating clinical validation. The clean tabular layout enhances readability and supports efficient data review in real-world diagnostic settings.



	CZB_ID	sequencing_batch	gender	age	SC2_PCR	SC2_rpm	idseq_sample_name
0	RR057e_00202	SEQ005	F	62.0	NEG	0.301622	RR057e_00202_N05_S78
1	RR057e_00080	SEQ003	M	81.0	NEG	0.091199	RR057e_00080_H20_S312
2	RR057e_00287	SEQ005	F	76.0	NEG	0.763690	RR057e_00287_L09_S140
3	RR057e_00753	SEQ005	F	36.0	POS	350.242314	RR057e_00753_G10_S151
4	RR057e_00751	SEQ005	F	58.0	POS	96314.765870	RR057e_00751_C10_S147
5	RR057e_00520	SEQ005	F	79.0	NEG	0.394572	RR057e_00520_I04_S57
6	RR057e_00505	SEQ005	M	89.0	NEG	0.249940	RR057e_00505_K04_S59
7	RR057l_00083	SEQ003	F	63.0	NEG	0.117495	RR057l_00083_K09_S139
8	RR057e_00764	SEQ003	F	71.0	NEG	0.076102	RR057e_00764_F16_S246
9	RR057e_00451	SEQ003	M	78.0	NEG	0.000000	RR057e_00451_I10_S153
10	RR057e_00285	SEQ005	F	83.0	NEG	0.300967	RR057e_00285_F11_S166
11	RR057e_00254	SEQ005	M	45.0	NEG	0.535249	RR057e_00254_J11_S170
12	RR057e_00726	SEQ003	M	74.0	POS	1056.448745	RR057e_00726_J06_S90
13	RR057e_00060	SEQ005	M	73.0	NEG	0.268898	RR057e_00060_F03_S38
14	RR057e_00560	SEQ005	M	89.0	NEG	0.794893	RR057e_00560_C06_S83
15	RR057e_00755	SEQ005	F	69.0	POS	6.006663	RR057e_00755_I10_S153
16	RR057e_00165	SEQ005	M	56.0	NEG	0.352707	RR057e_00165_L07_S108
17	RR057e_00399	SEQ003	M	73.0	NEG	0.102108	RR057e_00399_C04_S51
18	RR057l_00157	SEQ003	F	76.0	POS	770097.174900	RR057l_00157_D15_S228
19	RR057l_00167	SEQ003	M	71.0	POS	270.108609	RR057l_00167_L17_S268

Fig. 7: User Interface visualizing the Prediction results can be downloaded

Table 1 presents a comprehensive performance comparison of eight classification models used for COVID-19 viral status prediction based on gene expression and clinical metadata. It displays four key evaluation metrics Accuracy (A), Precision (P), Recall (R), and F1-Score (F1)—for each classifier in a structured, tabular format. The models are listed in ascending order of accuracy, ranging from K-Nearest Neighbors (lowest at 0.8156) to the Proposed Hybrid DL model (highest at 1.0000). Gradient Boosting achieves a strong second-place performance with 0.9716 across all metrics, while traditional models like SVM, Decision Tree, and AdaBoost show moderate and identical results around 0.8369. This table effectively highlights the superior predictive capability of ensemble and deep learning approaches, particularly the flawless performance of the Proposed Hybrid DL classifier.

Table 1: Performance comparison of all Classifier models.

Classifier	Accuracy (A)	Precision (P)	Recall (R)	F1-Score (F1)
K-Nearest Neighbors	0.8156	0.7997	0.8156	0.8038
Support Vector Machine	0.8369	0.8213	0.8369	0.8110
Decision Tree	0.8369	0.8213	0.8369	0.8110
AdaBoost	0.8369	0.8213	0.8369	0.8110
Logistic Regression	0.8440	0.8528	0.8440	0.8031
Random Forest	0.9291	0.9392	0.9291	0.9229
Gradient Boosting	0.9716	0.9734	0.9716	0.9709
Proposed Hybrid DL	1.0000	1.0000	1.0000	1.0000

5. CONCLUSION

The proposed DGSC for COVID-19 Strains presents a powerful end-to-end hybrid deep learning framework that accurately distinguishes between genomic samples of COVID-19 (SC2), other viruses,

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

and non-viral sequences. By combining deep neural architectures with advanced ensemble mechanisms, the model efficiently captures complex genomic signatures and ensures high discriminative power across diverse viral patterns. The comparative performance analysis across multiple classifiers reveals those existing algorithms such as K-Nearest Neighbors (Accuracy: 0.8156, F1: 0.8038), Support Vector Machine (Accuracy: 0.8369, F1: 0.8110), Decision Tree (Accuracy: 0.8369, F1: 0.8110), AdaBoost (Accuracy: 0.8369, F1: 0.8110), and Logistic Regression (Accuracy: 0.8440, F1: 0.8031) exhibit moderate predictive strength. In contrast, ensemble models like Random Forest (Accuracy: 0.9291, F1: 0.9229) and Gradient Boosting (Accuracy: 0.9716, F1: 0.9709) significantly enhance classification stability and reduce variance. However, the Proposed Hybrid Deep Learning model remarkably achieves perfect performance with Accuracy = 1.0000, Precision = 1.0000, Recall = 1.0000, and F1-Score = 1.0000, confirming its robustness and superior generalization. The system's integration of deep feature extraction with interpretable ensemble decision layers enhances both precision and explainability, making it a promising diagnostic framework. The model demonstrates excellent reliability for genomic-level COVID-19 detection, offering significant potential for scalable implementation in genomic surveillance, biomedical diagnostics, and epidemiological forecasting, contributing to more accurate, automated, and interpretable healthcare intelligence solutions.

REFERENCES

[1] Johns Hopkins University COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU). Available online: https://coronavirus.jhu.edu/map.html .

[2] Huang, C.; Wang, Y.; Li, X.; Ren, L.; Zhao, J.; Hu, Y.; Zhang, L.; Fan, G.; Xu, J.; Gu, X.; et al. Clinical Features of Patients Infected with 2019 Novel Coronavirus in Wuhan, China. Lancet 2020, 395, 497–506.

[3] Buitrago-Garcia, D.; Egli-Gany, D.; Counotte, M.J.; Hossmann, S.; Imeri, H.; Ipekci, A.M.; Salanti, G.; Low, N. Occurrence and Transmission Potential of Asymptomatic and Presymptomatic SARSCoV-2 Infections: A Living Systematic Review and Meta-Analysis. PLoS Med. 2020, 17, e1003346.

[4] Zsichla, L.; Müller, V. Risk Factors of Severe COVID-19: A Review of Host, Viral and Environmental Factors. Viruses 2023, 15, 175.

[5] Niemi, M.E.K.; Karjalainen, J.; Liao, R.G.; Neale, B.M.; Daly, M.; Ganna, A.; Pathak, G.A.; Andrews, S.J.; Kanai, M.; Veerapen, K.; et al. Mapping the Human Genetic Architecture of COVID-19. Nature 2021, 600, 472–477.

[6] Mutawa, A.M. Attention-Based Hybrid Deep Learning Models for Classifying COVID-19 Genome Sequences. AI 2025, 6, 4. https://doi.org/10.3390/ai6010004

[7] Deida, J.; Papa Mze, N.; Beye, M.; Ahmed, S.M.; El Bara, A.; Bollahi, M.A.; Basco, L.; Ould Mohamed Salem Boukhary, A.; Fournier, P.-E. Genomic Characterization of SARS-CoV-2 Variants from Clinical Isolates during the COVID-19 Epidemic in Mauritania. Genes 2024, 15, 361. https://doi.org/10.3390/genes1503036

[8] Azevedo, K.S., de Souza, L.C., Coutinho, M.G.F. et al. Deepvirusclassifier: a deep learning tool for classifying SARS-CoV-2 based on viral subtypes within the coronaviridae family. BMC Bioinformatics 25, 231 (2024). https://doi.org/10.1186/s12859-024-05754-1

[9] Nirmalarajah, K., Aftanas, P., Barati, S. et al. Identification of patient demographic, clinical, and SARS-CoV-2 genomic factors associated with severe COVID-19 using supervised machine learning: a retrospective multicenter study. BMC Infect Dis 25, 132 (2025). https://doi.org/10.1186/s12879-025-10450-3

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

[10] Aruna Rajalingam, Chaitra Mallasandra Krishnappa, Shanker Govindaswamy, Anjali Ganjiwale, Identification of Autoantigen Markers for SARS-CoV-2 Infection with Machine Learning-based Feature Selection: An Insight into COVID Symptoms, Coronaviruses; Volume 6, Issue 2, Year 2025, e250324228309. DOI: 10.2174/0126667975296293240320041641

[11] R. Pandiarajan and M. G, "Applications of Machine Learning in SARS-COV-2 Genomics for Varient Detection and Disease Prediction," 2024 International Conference on IT Innovation and Knowledge Discovery (ITIKD), Manama, Bahrain, 2025, pp. 1-6, doi: 10.1109/ITIKD63574.2025.11004732.

[12] Dubey, Shivendra & Verma, Dinesh & Kumar, Mahesh. (2025). Identification of Infectious Disease Like COVID-19 Gene Biomarkers Using a Clear Artificial Intelligence Approach. SN Computer Science. 6. 10.1007/s42979-025-03805-9.