Predicting Infection Positivity, Risk Estimation, and Disease Prognosis in Dengue Infected Patients by ML Expert System

Supreet Kaur 1, Sandeep Sharma 1, Ateeq Ur Rehman 2, Elsayed Tag Eldin 3, Nivin A. Ghamry 4, Muhammad Shafiq 5,*, and Salil Bharany 1,*

1 Department of Computer Engineering and Technology, Guru Nanak Dev University, Amritsar 143005, India
2 Department of Electrical Engineering, Government College University, Lahore 54000, Pakistan
3 Faculty of Engineering and Technology, Future University in Egypt, New Cairo 11835, Egypt
4 Faculty of Computers and Artificial intelligence, Cairo University, Giza 12613, Egypt
5 Department of Information and Communication Engineering, Yeungnam University, Gyeongsan 38541, Korea
* Correspondence: shafiq@ynu.ac.kr (M.S.); salil.bharany@gmail.com (S.B.)

Abstract: Dengue fever has earned the title of a rapidly growing global epidemic since the disease-causing mosquito has adapted to colder countries, breaking the notion of dengue being a tropical/subtropical disease only. This infectious time bomb demands timely and proper treatment as it affects vital body functions, often resulting in multiple organ failures once thrombocytopenia and internal bleeding manifest in the patients, adding to morbidity and mortality. In this paper, a tool is used for data collection and analysis for predicting dengue infection presence and estimating risk levels to identify which group of dengue infections the patient suffers from, using a machine-learning-based tertiary classification technique. Based on symptomatic and clinical investigations, the system performs real-time diagnosis. It uses warning indicators to alert the patient of possible internal hemorrhage, warning them to seek medical assistance in case of this disease-related emergency. The proposed model predicts infection levels in a patient based on the classification provided by the World Health Organization, i.e., dengue fever, dengue hemorrhagic fever, and dengue shock syndrome, acquiring considerably high accuracy of over 90% along with high sensitivity and specificity values. The experimental evaluation of the proposed model acknowledges performance efficiency and utilization through statistical approaches.

Keywords: epidemic; dengue disease; machine learning; prediction; risk level forecasting; tertiary classification

1. Introduction
The burden of vector-borne diseases is highest in tropical and subtropical areas. It accounts for more than 17% of all infectious diseases globally, resulting in a death toll of more than 700,000 annually. The term “vector” is synonymous with a mosquito-borne illness, as vectors are the living organisms that can transmit infectious diseases caused by bacteria, viruses, and parasites between humans or from animals to humans. Among the global top deadliest diseases, yellow fever, Zika fever, Chikungunya, and dengue fever are transmitted mainly by the mosquito’s genus, Aedes Aegypti. The female Aedes aegypti mosquito feeds on humans and is responsible for spreading the highly infectious dengue disease. A global pandemic of dengue began after World War II (in southeast Asia) and intensified over the next 15 years, making it a rapidly emerging pandemic disease in tropical and subtropical regions of the globe [1]. The distribution of vector-borne diseases are determined by complex demographic and social factors as pathogen transmission is highly impacted by the change in global travel/trade, unplanned urbanization, and environmental challenges such as climate change, change in agricultural practices, variation in temperature and rainfall, growth of urban slums, lack of reliable piped water, inadequate solid waste management, and many other factors [2]. This is why the only sure way of
combating mosquito-borne disease is the complete eradication of mosquitoes, which may sound feasible and easy theoretically, but in practicality is murky. In its early stages, dengue presents a ‘flu-like’ syndrome because it is often misdiagnosed or, in most cases, neglected. Among many diseases, Chikungunya, malaria, tick and epidemic typhus, enteric fever, and meningococemia conditions that exhibit similar symptoms to dengue clinically present another challenge in dealing with this disease. The time between exposure and the onset of symptoms ranges from 2 to 14 days, but most often, it is 4–7 days. Hence, travelers returning from epidemic zones are likely to have dengue if fever or other symptoms start within 3–14 days after reaching their homeland. WHO [3] divides dengue infections into three groups based on how bad they are: DF (dengue fever), DHF (dengue hemorrhagic fever), and DSS (dengue shock syndrome).

The researchers further studied that the whole infection cycle’s survivability increases as the temperature increases and concluded that they can hardly withstand cold temperatures, inferring from periodic incidences where the dengue epidemic always peaks near the rainy season and rarely prevails when the next season, i.e., the winter season, is in full force. However, scientists have now also found traces of dengue infection disease in those parts of the world which experience the coldest climatic conditions throughout, hence highlighting the threat of adaptability of this infection-carrying mosquito and at the same time negating the previous claim that dengue is a disease of tropical or sub-tropical countries [3]. Each year, dengue claims more than 390 million infections per year, as recent estimations put half of the world’s population at risk. The current COVID-19 pandemic has already told the whole world how insufficient human efforts have become to timely curb the issue when a disease escalates to a pandemic level. The global research teams came together in a race against the COVID-19 pandemic. Many machine learning and deep learning algorithms were trained, using any available data to help medical staff save lives.

Similarly, applying a supervised machine learning algorithm allows the learning program to learn from hidden patterns in the input dengue dataset and use this experience while classifying a new observation [4]. Moreover, with the inclusion of IoT devices and cloud computing, the availability of real-time data has increased tenfold, enabling a real-time disease management system that can save numerous lives by making needed interventions at the earliest possible predicted warning stage. There are various popular classifiers in machine learning, such as logistic regression, naive Bayes, k-nearest neighbor, support vector machine, random forest, and decision tree. To ensure the best working of the proposed model consisting of a tree-based classifier, various conditions have been performed to improve prediction, such as eliminating duplicate data and missing value data, using label encoders, and applying stratified sampling to conserve the percentage of sample for each resultant class, along with the inclusion of more than 17 attributes, which were restricted to 13 in previous studies.

The summary of the main contributions of this study can be broadly categorized as follows:

- To emphasize the need to predict dengue infection effectively and precisely.
- To propose a tool for data collection and analysis for dengue patients.
- To provide a real-time patient diagnosis of dengue infection present or not by analyzing more than 17 data attributes.
- To further predict if the identified dengue-infected patient has a pattern of a possible internal hemorrhagic manifestation that may threaten life as a severe case of thrombocytopenia.
- To examine the proposed model’s efficiency in identifying which stage of dengue infection level the positive dengue patient may be suffering i.e., DHF or DSS.
- To provide alerts to concerned stakeholders for providing timely medical aid.

The rest of the article is structured as follows, Section 2 “Related work” summarizes the current related work. The model to predict dengue-infected patients is proposed in Section 3 “The Proposed Model”. The model’s experimentation results and performance analysis are conducted in “Results and Discussions”. Finally, Section 5 “Conclusions” gives the conclusions derived from the proposed model in this study.
2. Related Work

2.1. Dengue Fever Infection

The life cycle of dengue infection transmission is illustrated in Figure 1, which follows stages: the first being where the adult female Aedes (whose body, as well as legs, are black, and who has distinctive white patches throughout (often called tiger strips)), after having a blood meal (rich in protein), lays black oval-shaped eggs over a period of three days above clean water surfaces, whose viability is very high due to the presence of chorion ranging between six months to a year. After just one or two days, the eggs proceed to the next feeding stage, i.e., larvae, which rest at an angle and have a short and stout syphon with one pair of hair tufts to the clean and non-polluted water surface. After a span of six to eight days, larvae go into a non-feeding stage known as pupa, having a long breeding trumpet with a narrow and slender opening from which, after one to two days, the next stage, i.e., adult, emerges, which can further start many other mosquito life cycles by laying eggs after a successful blood meal.

![Figure 1. Dengue infection transmission cycle.](image-url)

Usually, people infected with dengue virus are asymptomatic, having only mild symptoms such as fever. In contrast, some may have a further severe illness that is, in rare cases, also life-threatening. It affects numerous bodily systems (either in isolation or along with the classic dengue symptoms), which may lead to a decreased level of consciousness attributable to either inflammation of the brain or not directly as a consequence of impairment of vital organs, such as the liver. The only way to identify the infected person and his exact condition is by making that person undergo several clinical tests. However, all these clinical tests may not be globally available at an affordable price. However, the ever-escalating number of people suffering from this disease and limited healthcare re-
sources in certain parts of the world may result in inaccurate diagnoses and, consequently, wrong treatment plans [9,10]. This all may eventually lead to a sudden increase in the mortality of patients, thereby exponentially affecting the rapid transmission of infection, which ultimately adds to a big challenge for the healthcare sector. It may be concluded that more work must still be carried out on DF. Overall, very few researchers worked on correctly identifying infected patients and categorizing them based on severity level, such as DF, DHF, or DSS [11–13]. Even if they used a clinical dataset, i.e., patient’s clinical test results as primary input, they struggled to evaluate from which stage the patient was suffering or the prognosis of one stage to a more deadly one, and the main focus is only on identifying infected people among the population [14–16]. It was observed that studies including DHF and DSS had their main deciding parameters as bleeding and low platelet count only [17–20]. Still, in various sessions with expert doctors on this topic, they highlighted that though at the start, these are the main indicators, when the disease progresses into DHF or DSS various other clinical parameters also need to be monitored as by now multiple organ involvement is present which could result in multiple organ failure [21,22].

Most of the studies on dengue disease focused on the relationship between meteorological data and dengue deaths/cases, allowing researchers to forecast dengue infection case trends even up to six months in advance [23]. This helps establish different frameworks that could help medical professionals take timely preventive actions against dengue epidemiology. Some researched the vector aspect, i.e., mosquitoes and the human role in the dengue infection cycle. Including air-traffic data along with dengue, death/cases, and meteorological data could help navigate human mobility, which is the prime suspect of this widespread infection [24–27]. While analyzing the meteorological data, various parameters such as rainfall, temperature, and humidity are also used in predicting and controlling the dengue mosquito population by monitoring their breeding areas [25–27] to implement preventive approaches [27–32] and prevent the epidemic’s spread. All the proposed models have claimed to generate accuracy above 80% in most cases.

In contrast, some have been successful in forecasting outbreaks from months to years in advance. Risk estimation of dengue outbreaks in studies carried out in Malaysia, Thailand, and Singapore have shown high accuracy [28–31]. However, sites with few or no resources are a significant limitation that impedes the full functionality of such systems; additionally, the lack of parametric inclusion about dengue symptoms and the patient’s key health aspects cause the overall approach to fall short of total exposure [32,33]. A newer method includes proposed systems using sensors to monitor the mosquito population to raise a warning to the dedicated government bodies for possible risk areas; however, its implementation will evaluate the approach’s efficiency [34].

2.2. Cloud Computing and IoT-Based Dengue Healthcare

Including other factors such as cloud computing and IoT gives the dengue predictive approach a cutting-edge advantage over its peers, as the real-time data makes the model’s work more efficient [35–37]. IoT-based sensors and healthcare systems have made the dengue-combating system more resilient, which not only detects infected patients but also, on admittance to a hospital, helps in disease management [38,39]. If included in IoT systems [40,41], artificial intelligence can easily handle the data generated by the sensors, which need the supervision of medical personnel for understanding. Moreover, these techniques effectively detect any anomaly, enhancing the security aspect of IoT-based healthcare systems and making them intelligent [42,43]. A wearable bio-sensor gives real-time data, which can be used as a data input point and, in warning times, could alert the person instantly [44]. A fully capable IoT-based healthcare center is proposed as the framework which can effectively handle an influx of patients with ease in minimal time, and can also be a great aid in monitoring patients in the critical stage, by alarming them about any danger, even in the absence of medical staff [45]. Using the same approach, the researchers used a regularization genome-wide association study to predict colorectal cancer. They gained a 96% precision rate, the highest in comparison to SVM and semiconductor nano-
structured [46]. Similarly, in another study, the system generated warnings by SMS and emails once they received a confirmed diagnosis and treatment from the doctor [47]. Although the studies on IoT sensor-based healthcare claimed the proposed system to be more efficient and effective than the existing ones, the inclusion of the traditional security aspect in the proposed models becomes a major setback against hackers, and a new cryptography mechanism has been suggested to enhance the security of IoT-based healthcare systems [48,49].

3. The Proposed Model

We propose a model for dengue data collection and diagnosis for dengue-infected patients. The model’s core lies in dengue infection prediction, which works in two tiers; in tier 1, the system works to classify dengue-infected persons with warning signs from those without warning signs and non-infected persons and classifies persons into three classes, i.e., no dengue, dengue fever, and dengue fever with warning signs. The main emphasis in tier 2 is on predicting which level of dengue infection the infected patient is suffering from. This classifies infected patients into three classes: dengue fever with no bleeding, dengue hemorrhagic fever, and dengue shock syndrome. Figure 2 shows WHO’s classification of this infection into three levels: DF, DHF, and DSS, with each being more deadly than its predecessor, and also shows a set of conditions for each level. These were the criteria followed throughout this study for dengue predictions. Before proceeding to the overall workflow of the proposed algorithm, the next part of this paper deals with the basic architecture and explanation of the two-tier dengue diagnosis approach.

**Figure 2.** Latest WHO dengue disease classification.
3.1. Model Architecture

The proposed architecture consists of three stages, namely data collection, prediction technique, and validation. Figure 3 explains the diagram and data flow for this phase of the proposed system to give tertiary classification. However, the proposed model overall works in two phases. In the first phase, using only symptomatic data, the model makes the predictions, i.e., among the input data, the model works to find whether the target person has dengue fever (DF), no dengue fever (ND), or dengue fever with warning signs (DFW), and if the result is either of DF or DFW, a further clinical investigation begins. The proposed system on encountering any warning signs sends alerts to the registered user’s device, warning them of possible internal hemorrhage and thrombocytopenia.

Figure 3. Proposed system phase-1 architecture.

Phase 1. Dengue infection Prediction

- Data Collection: The dengue patient information was screened and retrieved for those admitted/referred to the government medical college in Amritsar for dengue monitoring as well as those who were dengue positive but hospitalization was not required. The positive patients were confirmed to have dengue infection through NS1 antigen or IgM/IgG tests. The vital signs and clinical test parameters were then recorded in a Microsoft Excel sheet for further preprocessing, in which missing value records were dropped. This study works with 19 data attributes that, to the best of the author’s knowledge, have not been included in only the symptomatic dengue infection model.
- Prediction Model: The random forest tree algorithm in the machine learning library of Python was trained by using training data, but before that, the stratified k-fold cross-
Validation for the value of $k = 6$ was applied to the pre-processed dataset generated in the previous phase to obtain quality random training and testing dataset. The predicting algorithms were saved in the Jupyter Notebook for further work.

- **Validation and Improvement:** The predictive models were tested using a testing dataset generated in Phase 2, and the accuracy for each was calculated. The confusion matrix was then made to check how well the models work, and the values of precision, F1 score, sensitivity, and specificity were calculated so that the models can be compared.

In the second phase, as shown in Figure 4, the proposed model works on patients who became positive in previous predictions, i.e., dengue fever and dengue fever with a warning. Using clinical data, the model makes predictions about the level of infection present in the target person as per WHO classification, i.e., DHF and DSS, to elaborate risk levels. Nevertheless, during expert sessions, it was found that there may be dengue fever with indicators that may seem to be DHF but do not always come under DHF or DSS classification. Therefore, in this phase, the proposed system predicts whether a dengue-positive patient is suffering from dengue fever with no hemorrhage, dengue hemorrhage fever, or dengue shock syndrome. The mortality rate increases as the disease progresses from DF to DHF and is ultimately life-threatening with high mortality.

![Figure 4. Proposed System Phase-2 Architecture.](image)

**Phase 2. Prediction of Dengue Infection Levels**

- **Data Collection:** The dengue patient’s positive information was screened based on NS1 antigen or IgM/IgG test and retrieved from the government medical college in Amritsar. The clinical test parameters were extracted from their file and then were recorded in a Microsoft Excel sheet for further preprocessing in which missing value rows were dropped even if one column was null.

- **Prediction Model:** The stratified k-fold cross-validation for the value of $k = 8$ was applied to the dataset generated, which was further inputted to the random forest tree classifier. The predicting algorithms were saved in Jupyter Notebook for further work.
Validation and Improvement: The predictive model was tested using a testing dataset, and the model's accuracy was calculated. The confusion matrix is then made to check the performance of the models based on the values for precision, F1 score, sensitivity, and specificity.

3.2. Proposed Dengue Diagnosis System’s Workflow

During the study period, the major obstacle observed was the data to work with, which was not readily available, in contrast to dengue death and report cases and a metrological dataset that could easily be accessed through government health sector websites. As discussed in the later section of this paper, the dataset required for a symptomatic investigation study was collected manually from hospitals, health centers, etc. The data available at these centers was in hard copies, which further requires skills to preprocess and clean the data before processing it into soft files. Therefore, in this paper, a tool for data collection and analysis for dengue patients is proposed, which will be web-based in Django using MySQL as the backend and Python for data analysis, including data determinants such as NumPy and pandas, etc. The proposed system’s data collections have three access modes: patient data by self-reporting, hospital interface for doctors and medical professionals, and an Excel import utility.

Similarly, the proposed system has three access interfaces: patient self-reporting, hospital interface, and admin interface. The hospital interface is available to all the partnered hospitals; a list of these hospitals is also available on the system, which can be referred to by the user when in need. In Table 1, the user interface workflow is discussed, in which once the user is registered with the system, he/she posts his/her symptoms and clinical test results and receives predictions accordingly, after following various steps.

Table 1. System workflow for dengue diagnosis model.

<table>
<thead>
<tr>
<th>Step No.</th>
<th>Suggestions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 1</td>
<td>Each user registers with the system by using a mobile phone or website. Identification health ID is automatically generated by the system and allotted to each user after registration (each health ID generated can further include five distinct health records, which may be family or friends).</td>
</tr>
<tr>
<td>Step 2</td>
<td>During registration, the user needs to share information about their name, age, sex, mobile number, address, UID, and state, before proceeding to the next stage.</td>
</tr>
</tbody>
</table>
| Step 3   | A form appears in which all possible dengue disease symptoms are mentioned and the user is urged to select symptoms applicable to their present condition, along with the duration they have experienced those symptoms (in days).  
  - Only if the user is female, unusual heavy menstruation bleeding and pregnancy options are visible, which otherwise are hidden. |
| Step 4   | The random forest algorithm is used to classify the user’s category into infected, not infected, or infected with warning signs. |
| Step 5   | Results are shown and based on the symptomatic period experienced. Dengue confirmation test is also advised for the infected user as follows:  
  - If days are less than 5 then NS1 antigen dengue test  
  - If days more than 5 then IgM/IgG dengue test. |
| Step 6   | If the result shows dengue infection with warning signs, then based on symptom mining, an alert message is sent to advise additional clinical tests which may include:  
  - Renal profile  
  - Liver profile  
  - Hematology—Hb, TLC, DLC, platelet count. |
| Step 7   | The infected user comes back and updates records to input all the test’s parametric values after undergoing tests recommended by the system, and this goes again into a random forest algorithm to generate a classification of dengue infection level. |
| Step 8   | The system classifies the user’s category as DFNH (dengue fever with no hemorrhage), DHF (dengue hemorrhage fever), or DSS (dengue shock syndrome), and messages alerting the user to seek medical aid are are sent accordingly. |
| Step 9   | Users can update information in their records anytime, collected from wearable sensors, test reports, etc. |
| Step 10  | The classification metrics are calculated to examine the accuracy of the proposed system and at each diagnosis, the user can print the complete decision, set i.e., the combination set of symptoms observed or percentage certainty of acquiring the disease according to the analysis, which can be shown to a doctor to further aid in receiving proper medical help. |
3.3. Data Collection

The data required for this study has been collected manually over three years from different districts of the Punjab state in India, to avoid stagnation of symptom pattern sets, as well as to have a diverse dengue infection symptoms trend, as suggested by the dengue nodal officer at GMC (Government Medical College) Amritsar during various expert discussions held. The doctor informed us that in each dengue active season yearly, certain combinations of symptoms are repetitive: for example, in the year 2018 every other dengue patient had liver involvement, whereas in the year 2019 it was observed that retro-orbital headache with kidney organ involvement was prominent. Therefore, they advised us to include different years' data and that from different source cities, in order to study diverse trends. All ethical and legal formalities were duly taken care of before and during the research period. Data was present in the hard files of each person coming to the hospital with suspicion of dengue infection, which included details such as name, age, sex, contact number, symptoms observed, the onset of symptoms date, tests administered to the person, test reports including case reports about hospital stay, and a full medical log, if any patient was in a critical stage. We collected data on confirmed dengue infected patients and critical dengue infected patients, as well as people who were suspected of having dengue infection based on symptoms but testing dengue negative. Microsoft Excel software was used to input data collected so far and this was grouped according to three classes, namely; ND (no dengue), DF (dengue fever infection), and DFW (dengue fever with warning) for phase 1 of the proposed model's dataset. These dataset attributes are age, UID, OPD/IPD id, mobile number, sex, address (city), fever, loss of appetite, nausea, rash, extreme weakness, retro-orbital headache, joint/muscle pain, mucosal bleed, blood in stool, abdomen pain, unusually heavy menstrual bleeding, pregnancy, loose motion, and large BP drop. There were no privacy rights violations of patients, and personal information such as postal address, phone number, name, Adhaar number, etc. was not extracted from hospital records for this study. In total, there were 19 data attributes in the input class and 3 data attributes in the output class. Only the label class had letters as values, i.e., DF, DFW, and ND, which were later converted into numeric values for operational ease. The dataset for symptomatic attributes was separated from the personal attributes data, which did not have any direct impact on the model's predictions. The clinical dataset file has attributes such as NS1 antigen, IgM, IgG, B.urea, S.Creatine, S.Bilirubin, SGOT, SGPT, TSP, DSP, Hb, TLC, platelet count, neutrophils, lymphocytes, monocytes, and eosinophils, which were gathered by referring patients that are dengue infected or were admitted to hospitals for having critical dengue infection and accounted for 17 data attributes. The three attributes of the dengue dataset, as discussed in previous statements, are depicted in Figure 5 as follows.

During the study for this model, it was found that researchers usually deal with only 12–13 data attributes from the symptomatic dataset, compared to around 7–8 for the clinical dataset. However, for this proposed model, 19 data attributes for symptomatic and 17 data attributes for clinical datasets were taken, which to date has never been considered, according to the best knowledge of the authors. Among 19 data attributes of the symptomatic dataset, unusual heavy menstrual cycle and pregnancy were considered for the first time in dengue disease research, along with others such as mucosal bleed, blood in stool, abdomen pain, bloody vomit, bloody urine, red spots, bloody nose, loose motion, large BP drop, etc. This allowed our proposed model to not only efficiently identify dengue-infected people, but also to raise an alarm for predicting possible hemorrhagic manifestations and thrombocytopenia without the use of any clinical test. Moreover, it was taken into account for the first time in the prediction that there could be cases of severe dengue fever (level 1 infection), which could be projected as DHF (level 2 infection) in the proposed model’s approach and labelled as dengue fever with no hemorrhage.
3.4. Train-Test Dataset

As previously discussed, some missing values and outliers, such as two patients who had a symptomatic period of one month that was out of the normal range, were excluded from this study. Furthermore, the fever attribute was not dropped or assumed to be always true for this study, as was observed to be common practice in previous research approaches, as we are also predicting a no-dengue class that includes patients who had some symptoms similar to dengue but were not dengue-infection positive. Moreover, the resultant classes were converted into numerical values for ease of computation as DF (0), DFW (1), and ND (2) for phase 1 and phase 2 as DFNH (0), DHF (1), and DSS (2), before processing using Jupyter Notebook. The 19 attributes encapsulating the patient’s symptoms were used as predictors against which the proposed model gives the final classification for phase 1, whereas for phase 2 there are 17 attributes. For phase 2, two columns were dropped to avoid redundancy because we are only considering patients’ records that are dengue positive, these included true/false for NG1 antigen test and IgM/IgG test results. A stratified k-fold cross-validation approach was applied and in the iteration process, different values of k were applied and k = 6 and k = 8 gave the best performance for phase 1 and phase 2, respectively, as shown in Figures 6 and 7. We used the stratified k-fold cross-validation technique to improve the quality of every iteration by ensuring that each data split was as good as the entire dataset, along with removing duplicate rows, which may have resulted in a smaller dataset than the original but improved the learning ability of the proposed model.
Figure 6. Data allocation (k = 6).
3.5. Classification

This module utilizes a random forest classification algorithm to classify infection and health risks of the registered user in real-time, and gives immediate medical test advice and alerts them to seek immediate medical support in case of a medical emergency related to dengue disease. Figure 8 shows a random forest classification tree in which dengue-related categorical attributes are inputted for the possible diagnosis of dengue infection. Based on the classifier’s output, if the user is dengue positive, the final health risk of the user is calculated and classified as no risk, when no strong evidence of infection is found; low risk, where the user has dengue fever but is not at any risk; and high risk, in which the user is dengue-infected but with warning signs.
The proposed classifier was trained using training data that had more than 1000 distinct symptom patterns, which were collected over three years of dengue active seasons to obtain a big chunk of quality data trends. During classification, it is often seen that fever symptoms are either dropped or taken as true [16–20] but as in this study tertiary classification was carried out, so this practice was not adopted. Contrary to popular belief, the proposed system also predicted patients who had symptoms similar to dengue but were actually dengue-negative. Furthermore, although symptoms such as pregnancy and unusually heavy menstrual bleeding are frequently overlooked in the dengue diagnosis dataset, we have included these as primary attributes if the registered user was female. One reason for this could be that among patients reported during the study, females comprised a very small quota in comparison to males, which is in line with previous research [15]. This study used a novel approach to estimate the risk levels of possible internal hemorrhage and thrombocytopenia in the absence of clinical tests. The inclusion of clinical test results has been standard practice in previous research, yet they only extract attributes such as Hb and platelet count most of the time [14,24–27]. However, experts negate this approach, as although thrombocytopenia is a distinct indicator of dengue disease, it alone cannot evaluate health risk of a person. To fully understand the infection risk level several other
attributes from clinical tests such as liver and renal profiles should be included to check for multiple organ involvement as the disease progresses. This paper has included not only these but also, on the advice of an expert doctor, Hb values were considered crucial data input [50–57]. Usually, previous researchers’ work has focused on low values of Hb and raised warnings for it, yet in the case of internal bleeding, which often happens in this disease, the value of Hb raises to as high as 20, which should have been normally 14. Therefore, with the inclusion of these new findings, a dataset is inputted into the classifier and based on the result, risk alerts are generated immediately to the concerned users [58–60].

3.6. Alert Generation

This module notifies the registered user about their diagnosis and health risk in real-time. When the user is dengue positive but is low-risk, they are advised to undergo a dengue clinical test and follow proper precautions; but when the user is dengue positive with warning signs, they are advised to seek medical help at the hospital on an urgent basis and their health record is also marked for continuous monitoring [61–65]. The partnered hospitals are also alerted about dengue-positive patients, especially those with warning signs [66,67] who need medical intensive care [68].

4. Experimental Evaluation

Machine learning algorithms were used to forecast the number of dengue cases based on a dataset of various variables. The results of the experiment are given in the section titled “results.” A tier 1 and a tier 2 set of findings have been generated through trials. This module discusses the experimental evaluation of the proposed system, including classification efficiency and alert generation efficiency. The efficiency of the model’s classification is very significant in achieving its objective. The current evaluation uses more than 2000 health records for training the random forest tree classifier using the Python toolbox. The training of the random forest classifier is iterated using a different number of epochs to minimize the error. The evaluation found that the error was minimized at the 50th epoch. After optimizing training, the classifier was tested with 7000 records at 50 epochs. Beginning with Figure 9a, we can see a comparison of the statistical performance of the proposed model with other employed classifiers for tier 1 using accuracy, sensitivity, specificity, F-measure, and precision. In Figure 10a–e, we compare the statistical performance of the proposed model to that of other employed classifiers for tier 2 predictions using accuracy, sensitivity, specificity, F-measure, and precision. The results acknowledge the proposed model not only has average high accuracy (94.16), but also sensitivity (0.90), specificity (1.00), f1-measure (0.94), and precision (1.00) in comparison to other classifiers employed in predicting dengue infected patients and patients with warning signs as targeted by tier 1 prediction models’ output. These statistical results proved the utility of a random forest tree classifier for dengue disease diagnosis, along with health risk classification. However, in the tier 2 predictions target, the model only achieved (88.80) accuracy, which was lower than some classifiers, yet the sensitivity, specificity, f1-measure, and precision of the proposed technique came close to (1.00), which represents its superiority over other classifiers employed in predicting dengue infection level when a classification metric was deployed for each. It clearly shows that the proposed model had better results in both predictions and better classification metrics, which validates further the selection of the random forest tree classifier among other selected machine learning classifiers for predicting dengue disease.
Figure 9. Statistical analyses (a) accuracy of different algorithms (b) sensitivity of different algorithms (c) specificity of different algorithms (d) F-measure of different algorithms (e) precision of different algorithms for tier 1 predictions.
Figure 10. Statistical analyses (a) precision of different algorithms (b) sensitivity of different algorithms (c) accuracy of different algorithms (d) F-measure of different algorithms (e) specificity of different algorithms for tier 2 predictions.
5. Conclusions

This paper proposes a model for dengue data collection and analysis to provide an accurate disease diagnosis. The proposed system addresses the issue concerning the effects of dengue infection on the human body and its vital organs through real-time diagnosis. The system works on two levels. Firstly, it predicts dengue-infected patients, among others, effectively only through symptomatic investigation, and raises the alarm whenever any warning trend is observed. Then, using clinical data, this system also aids in predicting dengue infection levels in the patient, which, according to WHO, are classified based on the increase in severity index as dengue fever, dengue hemorrhagic fever, and dengue shock syndrome. The proposed model alerts the registered users of warnings and urges them to seek medical support on an urgent basis. Not only this, but the model also suggests various clinical tests based on symptoms added by the registered user, which can further help correct disease diagnosis and consequently adopt a timely and better treatment plan. The predictions for dengue level analysis in the proposed model, unlike the traditional approach, does not only rely on parameters such as platelet count, Bp, and hemorrhaging but also further includes liver and renal profiles, as when disease prognosis from dengue fever reaches higher fatality levels, multiple organ involvement is often present, especially burdening the liver and kidney functions of the patient and leading towards organ failure. This proposed system mimics a real-world doctor’s behavior in analyzing and detecting dengue-related patients. Upon various expert suggestions, many new approaches were included for the first time. All aspects regarding dengue diagnosis, data collection, patient self-reporting, hospital reporting, and importing Excel files for cumulative data are organized and work incrementally in the proposed intelligent dengue disease diagnosis system. The experimentation evaluation of the system is based on health records of patients whose data has been collected over four years, and the classification metric evaluation of various modules justifies the utilization of the proposed model. The statistical results show the proposed random forest tree technique-based risk classification with other employed classifiers and its performance analysis; the alert generation performance acknowledges the performance efficiency and utilization of the proposed model. The proposed model in the current paper has considered the effect of dengue infection on vital organs such as the liver, kidney, and general hematology of the patient. Still, the system could be used to predict how a dengue infection will spread to other important organs, such as the heart, in the future.


Funding: This research was supported by the Future University Researchers Supporting Project Number FUESP-2020/48 at Future University in Egypt, New Cairo 11845, Egypt.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.
Abbreviations

DF Dengue fever  
DFNB Dengue fever with no bleeding  
DHF Dengue hemorrhagic fever  
DSS Dengue shock syndrome  
DFW Dengue fever with warning signs  
IOT Internet of Things  
ALT Alanine transaminase  
AST Aspartate aminotransferase

References

13. Pandiyarajan, P.; Thangairulappan, K. Classification of dengue serotypes using protein sequence based on rule extraction from neural network. In Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics); Springer: Berlin/Heidelberg, Germany, 2018; Volume 11308, pp. 127–137. [CrossRef]


47. Rehman, A.U.; Naqvi, R.A.; Rehman, A.; Paul, A.; Sadiq, M.T.; Hussain, D. A Trustworthy IoT Aware Mechanism as an Enabler for Citizen Services in Smart Cities. Electronics 2020, 9, 918. [CrossRef]


