REVIEW

BON VIEW PUBLISHING

Feature Selection, Clustering, and IoMT on Biomedical Engineering for COVID-19 Pandemic: A Comprehensive Review

Atikul Islam[1] , Soumita Seth[1,6] , Tapas Bhadra[1], Saurav Mallik[2,]*, Arup Roy[3] , Aimin Li[4] and Manash Sarkar[5]

[1]Department of Computer Science & Engineering, Aliah University, India

[2]Department of Environmental Health, Harvard T. H. Chan School of Public Health, USA

[3]School of Computing and Information Technology, Reva University, India

[4]School of Computer Science and Engineering, Xi'an University of Technology, China

[5]Department of Computer Science & Engineering, Atria Institute of Technology, India

[6]Department of Computer Science and Engineering, Brainware University, India

Abstract: In this era, feature clustering is a prominent technique in data mining. Feature clustering has also huge applications in biomedical research for multiple purposes including grouping, feature reduction, and many more. The Internet of Medical Things (IoMT) is a promising and emerging field of research that is having a major impact on knowledge retrieval and networking. IoMT also has significant application in biomedical research regarding remote monitoring and remote healthcare services. In this COVID-19 pandemic situation, psychological effects and human reactions have become a major concern of further research. A dataset can be reduced in size by using feature selection techniques. To facilitate subsequent processing, this will make the data easier to manage. Feature selection is also used to clean, reduce, and reduce dimensions of data. The clustering method has proven to be a powerful tool for finding patterns and structures in both labeled and unlabeled datasets. Our study basically provides various state-of-the-art methods regarding medical IoMT for remote healthcare, feature clustering for information retrieval regarding biomedical science. In this study, we have studied five different types of feature selection algorithms such as minimum redundancy maximum relevance (mRMR), random forest, normalized mutual information feature selection (NMIFS), F-test, and chi-square and five different types of clustering algorithms like hierarchical clustering, density-based spatial clustering of applications with noise (DBSCAN) clustering, K-means clustering, shrinkage clustering, and fuzzy C-means clustering. Finally, this study is very useful to understand and apply the appropriate IoMT, feature clustering, and catharsis on the various biomedical applications for the benevolence of society.

Keywords: IoMT, feature selection, clustering algorithm, COVID-19, biomedical engineering

Foreword

We are delighted to introduce this important and insightful study that explores the COVID-19 pandemic, IoMT, and feature clustering in biomedical engineering. It has been crucial during the COVID-19 pandemic to leverage the Internet of Medical Things (IoMT) for collecting health-related data in real time from connected medical devices and applications. IoMT enables these devices to be connected to the Internet.

The concept of feature clustering is important in biomedical engineering because it facilitates the identification of patterns and relationships between data features or attributes. Data from patient records, clinical trials, and epidemiological studies can be analyzed using feature clustering in the context of the COVID-19 pandemic. Diagnoses, treatments, and prevention of COVID-19 can be made easier by identifying meaningful patterns.

1. Introduction

In this manuscript, we have explained how bioinformatics technology can be applied in a COVID-19 scenario. The IoMT, a feature selection approach, and a clustering algorithm are examples of such technologies. It is critical to conduct patient monitoring, patient observation, dataset analysis, dataset modification, and other procedures during the COVID-19 pandemic.

IoMT, which stands for the IoMT, ought to be described another function with different foundations, concepts, appliances, methods, alongside with ideas from other accredited Internet access, in particular for the domains of medicinal drug and well-being program. IoMT treats orthopedics patients in a variety of ways, whether it is for bones, muscles, tendons, joints, ligaments, and soon. All orthopedic sufferers are faced with several integral and serious

*Corresponding author: Saurav Mallik, Department of Environmental Health, Harvard T. H. Chan School of Public Health, USA. Email: smallik@hsph.harvard.edu

troubles in such a harsh COVID-19 pandemic. IoMT deals with solutions along with redress for the above-mentioned orthopedic patient-related troubles by the use of advanced applied sciences and shrewd machine learning-based methods that have been introduced collectively to deliver fruitful proposals for the cure of orthopedic patients, especially in trendy COVID-19 pandemic background [1–3]. Modern datasets are very prosperous in facts with records gathered from hundreds of thousands of IoMT units as well as sensors.

This entails an excessive amount of spatial data, it is quite frequent to see file sets with lots of functions, and it is not outstanding for them to be tens of thousands. A healthy lifestyle is essential for living a peaceful and successful life. A state of fitness, according to the World Health Organization (WHO), is one in which the body and mind are free of disease and affliction. In health care, illnesses and injuries are prevented, diagnosed, and treated to maintain or enhance health. Human error and patient harm have been common outcomes of fitness structures' manual management and protection of client demographics, scientific history, diagnostics, medication, billing, and drug inventory. The IoMT enables medical practitioners to diagnose diseases more accurately and easily by connecting all essential symptom-monitoring devices together. IoMT refers to scientific matters that enable the exchange of information across networks without requiring human-to-human interaction [4]. In many societies, the aging population is increasing, emphasizing the importance of maintaining both intellectual and physical fitness for leading an independent life. Detecting, monitoring, and analyzing activities of daily living are essential steps to detect the fitness popularity of users, as they pave the way for providing immediate health services [5]. The intention of feature selection is to pick out relevant features enabling us to distinguish between patterns belonging to different classes [6]. The feature selection methods are generally selecting a subset of the unique characteristic set that represents the pleasant subgroup. The feature selection algorithm could be labeled in the direction of filters, wrappers, and embedded. The filter-based feature selection methods are commonly used as a preprocessing step. The selection of features is unbiased of the induction learning algorithm. Wrapper-based feature selection methods, on the other hand, utilize the getting-to-know classifier. Subsets are added or removed using this approach. Wrapper-based and filter-based feature selection methods are combined in embedded-based feature selection. Different inferences have been using for analyzing the excellence of a function such as statistics measures, analyzing distances, consistency measures, classification error measures, and dependency measures [7]. An algorithm of the decision tree is used in the construction of random forest, a supervised machine learning algorithm. It is used in numerous industries like e-commerce and banking and stock market outcomes and predictive behavior. This could be used for each classification and regression. It is given an effective way of controlling the missing data. It solves the over fitting trouble in the decision tree. The analysis of variance (ANOVA) *F*-test the purpose of feature enhancing the model. The aim of the chi-square feature selection test is to determine whether there is a significant association between two variables [8].

There are many feature selections used in bioinformatics engineering, most of which focus on data analysis [9]. Speech recognition, gene detection, and the diagnosis of different types tumors and cancers are included in the fields that use the mRMR feature selection. The purpose of the mRMR is to select features that can be correlated with the classification variable. The random forest feature selection is used in data science workflow.

The random forests improve the purity of the node. In order to determine redundancy with features, we use the normalized mutual information (NMI) [10].

The clustering algorithm is the determination behind the search engines. Search engines try to team similar objects in one cluster and the multiple objects from each other. Many clustering algorithms are used in this manuscript; we have discussed the shrinkage clustering algorithm. The shrinkage cluster is used to subtype cancer-based gene expression levels. The shrinkage clustering algorithm analyzes disease phenotypes hooked up on physiological and medical records [11]. Hierarchical clustering is a successful method that delivers us to bring together tree structures from data similarities [12]. The DBSCAN clustering algorithm is useful with unsupervised learning of the data. This clustering algorithm can discover outliers [13]. In fuzzy C-means clustering, each cluster is assigned a membership matrix, which determines the extent to which each sample is associated with the cluster [14]. Similarly, the K-means clustering algorithm is also useful for different datasets. This clustering algorithm is used to figure out crime-prone zone, public transport data analysis, customer segmentation, insurance fraud detection, document clustering, and clustering of IT alerts, etc. [15].

In the context of the COVID-19 pandemic, IoMT, feature selection, and clustering are essential components of patient treatment and monitoring. Biomedical engineering involves designing biomedical tools and systems to aid the recovery and improvement of individuals' health. In this article, the discussion on feature selection and clustering algorithms aims to enhance biomedical engineering [2].

2. Fundamental Term and Preliminary

This section defines and explains key concepts, theories, and terminology related to this manuscript. The purpose of this section is particularly important if your research involves specialized or technical terms not understood by all readers. A preliminary section explains the topic of this manuscript, its significance, and its context.

2.1. Supervised machine learning

The supervised machine learning is a subcategory of synthetic talent and machine learning. It is described through the use of tagged datasets to teach algorithms that categorize data or precisely outcomes. As input dataset is entered inside the supervised machine learning model, it adapted its weight just before the model is adjusted accordingly, whatever is section of the cross-validation process. The supervised machine learning helps organizations remedy a huge variety of real-world problems on a large scale. The goal of supervised machine learning is to acquire data in context to a unique query [16]. Many machine learning research activities have been focused on supervised learning. A wide range of data types can be examined and processed using supervised learning techniques [17–19].

2.2. Unsupervised machine learning

Unsupervised machine learning technique extracts patterns from independent data. The expectation is that imitation will pressure the machine to construct an impenetrable interior illustration of its world along with produce creative contented. Unsupervised machine learning solves the trouble with the aid of learning the data and classifying it except prior expertise of labels.

Analyzing data for patterns is very helpful. Visualization, dimension reduction, clustering, finding association rules, and anomaly detection are the core processes of unsupervised learning. Unsupervised learning tools are able to identify and select suitable features even when class labels are not desired, this is generally the case [20].

2.3. Mutual information

A mutual dependence is characterized by mutual information, which is delivered in the frequent distribution. Mutual information is one of the metrics used to quantify the degree to which one random variable reveals information about another. This dimensionless quantity, which is measured in bits (usually), may be conceptualized as a reduction in uncertainty regarding a particular random variable based on information about another. Mutual information is used to signify the maximum relevance and minimum redundancy of variables, such as the minimal redundancy feature selection [7]. Statistical dependence is measured by the mutual information. Previously, the idea was introduced. It has been suggested that mutual information should be used as a criterion for assessing independence [21].

2.4. Statistic test

Statistics and machine learning are two very carefully associated areas. These statistical methods can be used to cleanse data and put it together for modeling. These statistical speculation tests and estimation statistics can be useful resources in model selection and in presenting the skills and predictions of the ultimate models [18, 19]. There are several statistical tests that could be grouped into two classes, e.g., parametric and nonparametric tests. The parametric check is applied roughly which can have two subclasses unequal variance assumption and equal variance assumption. The famous statistical tests under equal variance assumption are Pearson's correlation test, *t*-test, and ANOVA (Anova 1) test [22]. Under the nonparametric test, the famous statistical tests are a linear model for microarray, significant evaluation of microarrays (SAM), Wilcoxon rank test, LIMMA test, etc. [23]. One common statistical test under unequal variance assumption is Welch's *t*-test. [24].

2.5. *T*-test

T-test is a peripheral method. In order to use a *t*-test, the samples must satisfy the conditions of normality, equal variance, and independence. A *t*-test can be performed on one sample, independent two-samples, or paired samples, depending on the nature of the data [25]. The abuse or improper usage of a statistical test may also cross the lookup in the incorrect approach and, as a result, flawed outcomes. The *t*-test is viewed as "the bread and butter" of statistical analysis [26]. It is important to take two factors into consideration when analyzing quantitative data effectively. In order to accurately assess whether the data fit the assumptions of the parameter tests, one must first determine what type of experimental design is being used [27].

2.6. ANOVA 1 test

ANOVA 1 is a statistical test used to compare means between two or more populations or groups. Through estimating variance estimate comparisons, we ought to create comparisons between group means. The variance is spitted into two parts, one is due to random error and other one is due to variations between the group's means. Later, the components of these variances are examined for statistical relevance. In the event that relevance holds true, we reject the null hypothesis which states that there is no difference in the population means as well as the alternative hypothesis indicating there is an important difference between each population. ANOVA is a statistical method in using scientific research. The most important endeavor of evaluation is centered on the versions of team means; however, ANOVA focuses on the distinction of variances [27, 28].

2.7. Filter method

Filter methods are typically applied as a step of preprocessing. The point's resolution is individual of any machine learning algorithm. Instead of choosing certain factors based on how well they ranked in multiple statistical analyses, consider their correlation with the outcome variable. For essential instruction, you could refer to the subsequent desk for describing correlation coefficients. The methodologies, in which features are selected independently of the classifier. are being employed [29, 30].

2.8. Pearson's correlation coefficient

Pearson's correlation coefficient refers to the correlation between two variables. This is used to determine whether there is a measurable linear dependency between the continuous variables x and y. These values range from -1 to 1. A point to keep in mind is that multi-co linearity is not eliminated by the filter method. Hence, you need to deal with multi-co linearity of elements as well beforehand education model of the data [31].
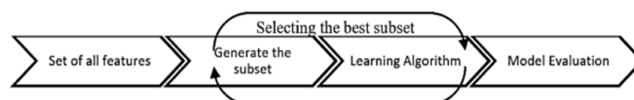
$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{(x_i - \bar{x}) \sum (y_i - \bar{y})^2}} \tag{1}$$

where r = Correlation coefficient, $x_i = x$ variable in a sample and x_i denote the value of sample, $\bar{x}$ = Mean of the values, $y_i = y$ variable in a sample and y_i denote the value of sample, and $\bar{y}$ = Mean of the values of the y variable.

2.9. Wrapper method

In wrapper methods (Figure 1), we try to practice feature subsets and educate a model through them. From the subset, we figure out to add or delete. These functions are usually computationally excessive cost. Some regularly occurring wrapper method examples are backward features elimination, recursive feature elimination, forward feature selection, and so forth [32]. A wrapper-based feature selection technique generates and calculates the prediction performance of a subset of features iteratively [33, 34].

Figure 1
Wrapper-based feature selection method

2.10. Backward features elimination

To improve the performance of model, we apply backward feature elimination, where we commence with the whole feature set and eliminate the features that are least significant at each iteration. We repeat until the removal of features shows no improvement. Backward feature elimination is employed to exclude variables that have a minimal impact on the output or prediction [32].

2.11. Embedded methods

The embedded feature selection methods or hybrid is a combination of filter and wrapping methods [29]. Wrapper and filter methods are combined to form embedded techniques. The most famous instance of these strategies is RIDGE and LASSO regression which have penalty functions to limit over fitting. Lasso regression employs L1 regularization, which imposes a penalty equal to the absolute value of the coefficient magnitude. Ridge regression applies penalization equal to the square of the magnitude of the coefficient through L2 regularization. [30].

2.12. COVID-19 pandemic

The pandemic situation was caused by uncontrolled spreading of corona virus that is known as COVID-19 Pandemic. During the month of December 2019, Wuhan, a Chinese city, declared an outbreak of the deadly disease. Feature selection is a critical component to handle the COVID-19 pandemic scenario. To find out the most relevant features for COVID and non-COVID predictions, feature selection (FS) approaches are used such as mutual information (MI) and Relief [35]. Also, clustering algorithms play important roles in doing so. The location, confirmed daily cases, recovery reports, and daily deaths are determined using clustering algorithms like C-means clustering and fuzzy C-means clustering [36].

3. Methodology

3.1. Method hierarchy
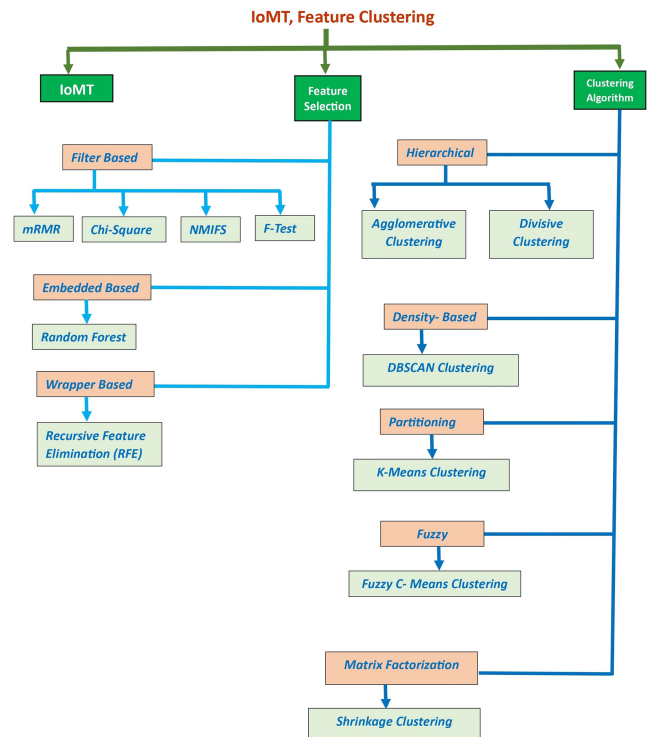
The method hierarchy is represented in Figure 2.

3.2. Description of methods

The current methodology for our work has been divided into three parts,
(a) **Internet of medical things (IoMT)** [2, 4].
(b) **Feature selections technique** [8]
(c) **Clustering algorithm** [9].

The Internet of things has been used in biomedical engineering for medical patient remedy and monitoring. Additionally, supervised and unsupervised features selection methods like mRMR (minimum redundancy maximum relevance) [37–42], random forest method [33, 43–45], NMIF(normalized mutual information feature selection) [21, 46], *F*-Test [47], and chi-square [47, 48] have been employed as a feature selection technique, and hierarchical clustering [49, 50], DBSCAN clustering [51, 52], shrinkage clustering [53], K-means clustering [54, 55], and fuzzy C-means clustering [56] have also been utilized as a supervised and unsupervised clustering algorithm.

Figure 2
Method hierarchy

3.3. IoMT

IoMT-based biomedical dimension structures play a crucial role in the analysis of numerous diseases such as heart attack, Genius tumor, Alzheimer's, Parkinson's, and epilepsy, as properly as extensively used for healthcare monitoring and treatment. During the COVID-19 pandemic, IoMT played a crucial role in the field of orthopedics, as highlighted by studies [2, 4].

(a) IoMT opens the door to interrupted fitness surveillance and presents proactive orthopedic care.
(b) IoMT creates an evolved tradition in the therapy method via the use of analytics along with latest digital devices; therefore, any feasible emergency may be estimated and analyzed remotely.
(c) IoMT proposal superior digital wearable gadgets for the appropriate patient monitoring such as pulse, variation of blood pressure, ECG measurement, etc., are checked and monitored.
(d) IoMT is a most important gain of fee reduction. IoMT can avoid charges that the patient incurs for popular doctor visits and tests.
(e) With this technology, wise orthopedic hospitals can be developed. As a result of the well-networked technology, a comprehensive hygiene monitoring system is maintained within the health facility, so that possible infections can be prevented and medical staff can treat patients quickly when necessary. The practice of real-time monitoring makes patient care more effective.
(f) Since the data are well monitored on a cloud basis, the real-time database assists in diagnosing diseases at a preliminary stage,

such as Early Detection of Cardiac Conditions, Remote Monitoring for Chronic Diseases, and Early Cancer Detection. Real-time databases serve as valuable tools for enhancing the diagnostic process and enabling proactive healthcare, but they are not standalone diagnostic tools.

(g) Since IoMT offers consistently with the connected system channel, the storage and consumption of medicinal drug can be calculated.

(h) The monitoring of a patient by a doctor can be intelligently guided by IoMT. The IoMT provides a top-notch facility in the situation of COVID-19, where it is impossible to engage in face-to-face interactions due to the pandemic.

3.4. Feature selections technique

As part of the predictive model development process, features are selected to reduce the number of input variables. The feature selection techniques to center of attention on the subset of relevant variables. Features are selected in three ways: using wrappers, filters, or embedded methods. Some feature selection methods are given below: [8, 57].

3.4.1. Minimal Redundancy Maximal Relevance (mRMR) Algorithm

This feature selection process prefers features with high correlations between themselves and with the class, a process known as the mRMR [42]. In most feature selection processes, features and classification classes are considered independently of each other, but the mutual influence between them is not considered. It is possible to weigh the information quantity between feature variables X and Y based on mutual information; the mRMR feature selection can measure how a great deal information X can furnish in order to also classify activities. As a result of these exceptional features, we can accelerate two standards: redundancy and relevance. The first part of the redundancy measure involves utilizing mutual statistics to quantify information. When the cost is large, a large amount of information is duplicated. Redundancy is a measure of feature selection criteria. It is preferable if the value is lower. [37–41, 59].

The mRMR is built on the concept that a function set 's' and 'm' represent aspects $\{X_i\}$ that are correlated with the type 'C' (output) and should be considered [58]. Maximum relevance is achieved by utilizing the suggested price of mutual information among all features $\{X_i\}$ with class C. The formal expression of two random variables X and Y mutually informing each other is as follows:

$$MI(X, Y) = \sum_{x \in X} \sum_{y \in Y} p(x, y) \log \frac{p(x, y)}{p(x)p(y)} \qquad (2)$$

Taking the variables in S and recognizing their relevance to C as a whole, we can calculate:

$$V_1(S) = \frac{1}{|S|} \sum_{X_i \in S} MI(C, X_i) \qquad (3)$$

Here, $MI(C, X_i)$ is the mutual information of feature X_i with class C.

Furthermore, mRMR proposes to supplement the maximum relevance criterion with a minimal redundancy criterion among features in order to enhance relevance. It is very probable that if the focus is solely on relevance, the dependence between features will also grow. So minimal redundancy desires to be utilized

barring demanding about its relevance. The following system desires to be used to get minimum redundancy between features.

$$W_1(S) = \frac{1}{|S|^2} \sum_{X_i, X_j \in S} MI(X_i, X_j) \qquad (4)$$

Here, $MI(X_i, X_j)$ is the mutual information of feature X_i with X_j.

mRMR refers to the criterion which combines the above two constraints. A desirable set of features can be acquired by optimizing relevance and redundancy in the following manner:

$$\max \varnothing (V_1(S), W_1(S)) \qquad (5)$$

where $\varnothing = (V_1(S) - W_1(S))$.

The mRMR feature selection algorithm is following steps give below: [38]

Step 1. In the original feature set Ω, the optimal feature x_i can be selected by $I(x_i; c)$ and after that put into the optimal feature subset S;

Step 2. In the feature subset $\Omega_s = \Omega - S$, the next optimal feature x_j is selected which satisfies (3) and (4);

Step 3. Repeat Step 2 to find the optimal feature subset S which meets the shaped requirement finally.

Advantage:

(1) mRMR feature selection is an algorithm often used in a technique to accurately find out traits of genes and phenotypes and slim down their relevance.

(2) "mRMR" is an algorithm used for discovering the "minimal-optimal" subset of features.

(3) mRMR feature selection has been located to be greater fine than the most relevance selection.

(4) mRMR tries to tackle this problem by casting off these redundant subsets.

(5) The mRMR algorithm represents a computationally greater possible approach.

Disadvantage:

(1) mRMR works iteratively. Once a characteristic goes into the bucket, it cannot ever come out.

(2) mRMR creates a data complexity in data processing.

(3) mRMR feature selection in the habitually used form consists of an immoderate sensitivity of massive relevance and redundancy measures to the presence of outliers in the data.

(4) mRMR is non-stop data that are discretized, which is actively attenuated due to a useless loss of information.

(5) The mRMR is no longer clear how to determine the threshold factor for rankings to select only the required features and eliminate noise

3.4.2. Random forest feature selection method

The random forest is one of the supervised feature selection methods. This method creates more than one selection trees and combine them collectively for an extra reliable and accurate prediction. In addition to its ability to use random forests to solve regression and classification problems, random forest feature selection has many other advantages. Let us recognize that the algorithm is used a sample dataset. Feature selection turns into prominent; it is particularly relevant to datasets that have a large

number of variables and features. Through this method, unimportant variables will be removed, resulting in an increase in accuracy consistent with how classification works on a typical basis. In spite of a greater quantity of variables, random forest has proved to be a fantastically helpful algorithm [43]. Nowadays, the random forest algorithms have been successfully applied to many scientific fields today in order to process large and multi-source datasets [33]. The random forest algorithm working process diagram is given below: [44, 45].

The random forest algorithm performs the process outlined in the steps. The steps are as follows: [43–45]

Step 1. To begin, select random samples from a dataset.
Step 2. As a next step, this algorithm builds a DT (decision tree) for every sample. Each decision tree should then produce a prediction result.
Step 3. During this step, sample datasets will be generated for every predicted result.
Step 4. Lastly, select the prediction result with the most datasets as the final prediction result.

Advantage:
(1) Random forest feature selection can be used to clear up both classifications as properly as regression problems.
(2) Random forest works well with each express and continual variable.
(3) Random forest feature selection can necessarily handle missing values.
(4) Random forest feature selection is typically strong to outliers and can manage them automatically.
(5) As a classification and regression tool, it can handle both types of problems.

Disadvantage:
(1) Random forest requires much more time to train.
(2) It also requires tons of time for training as it combines a lot of decision trees to decide the class.
(3) This algorithm requires tons of extra computational strength and resources.
(4) Random forest feature selection is comparatively low impacted by noise.
(5) Due to the ensemble of choice trees, it additionally suffers interpretability and fails to determine the value of each variable.

3.4.3. NMIFS

The Normalized Mutual Information Feature Selection is a supervised feature selection technique used to determine redundancy within features. The goal is to select features in order to be able to differentiate between patterns belonging to various groups. The NMIFS operates based on the mutual information between f_i and f_s, NI $(f_i; f_s)$ as the mutual information normalized by the minimum entropy of each feature, the feature selection problem normalized by using the condition [7, 21].

$$\text{NI } (f_i; f_s) = \frac{I(f_i; f_s)}{\min\{ H(f_i), H(f_s)\}} \quad (6)$$

This feature selection technique utilizes the NMI in order to determine of the redundancy between the subset of selected features $S = \{f_s\}$, *for* $s = 1, 2, 3, \ldots\ldots, S$

$$\frac{1}{|S|} = \sum_{f_s \in s} \text{NI } (f_i; f_s) \quad (7)$$

where $|S|$ is the cardinality of the set S and condition (7) is a type of correlation determine, that takes a value in [0,1]. A value of zero suggests no correlation between feature f_i and the subset '*S*'. A value of 1 indicates that all features in the subset S are highly correlated. Basically, NMIFS is a way of selecting features with the highest G measure, and this condition is given below:

$$G = I(C; f_i) - \frac{1}{|S|} \sum_{f_s \in s} \text{NI } (f_i; f_s) \quad (8)$$

The whole NMI feature selection algorithm proceeds in the following manner:
Step 1. Initialization: Set $F = \{f_i / i = 1, 2, \ldots\ldots, N\}$, A set of N feature is initially created and empty set.
Step 2. Determine $I(C; f_i)$ for each $f_i \in F$
Step 3. First feature: Find $f_i = max_{i=1,2,\ldots\ldots N} \{I(f_i; C)\}$. Set $\leftarrow F \{f_i\}$; *set* $S \leftarrow \{f_i\}$.
Step 4. Greedy selection: Repeat until $|s| = k$.
Step 5. The selected features are output in the set S.
Advantage:
(1) The concept underlying MI is to measure the mutual dependence between two random points with the aid of figuring out how much facts of one of the features can be derived from another feature.
(2) With NMI, MI costs were normalized to lower the price between zero and unity, removing the bias toward multivalued features.
(3) NMIFS to find the irrelevant and redundant features.
(4) NMI will increase the understanding of the classification with the aid of evaluating the quantity of records got from the clusters.
(5) The value of NMI is always between 0 and 1.

Disadvantage:
(1) This NMIFS is caused by a lack of statistics about the interactions between the facets and the classifier.
(2) There is a relationship between the redundancy in the features.
(4) Time period and the number of selected features when the variety of selected features increases in recognition of the relevancy term.
(5) As an inclusion criterion, NMIFS determines redundant or inappropriate functions and eliminates them.
(6) A measure of redundancy among features is proposed by using common NMI.
(7) Sequentially combining the NMI and fisher discriminated ratio.

3.4.4. ANOVA F-test feature selection

F-tests are statistical tests that are based on ANOVA. To find a mannequin that well matches the population from which the data samples are taken, statistical models are continuously used when analyzing datasets. To test if all extraordinarily different responses to Y have the same mean as X, a one-way ANOVA *F*-test was applied for each limitless predictor. The following mathematical expression is given [46, 60].

$$S_j^2 = \sum_{i=1}^{N_j} (x_{ij} - \bar{x}_j) / N(N_j - 1) \quad (9)$$

where N_j is a Cases involving $Y = j$.
$\bar{x}_j$ Is the mean of the sample predictors X for target class $Y = j$.

S_j^2 Is variance of the predictor sample X for target class $Y = j$.

$$\overline{x} = \sum_{j=1}^{j} \frac{N_j \overline{x}}{N} \qquad (10)$$

The above expressions are based on unmissable combination of (X, Y).

In that case, F-statistics is on the basis of p value, p value calculates by,

$$p\ value = Prob\ \{F(J-1,\ N-J) > F\} \qquad (11)$$

where

$$F = \frac{\sum_{j=1}^{J} N_j \left(\overline{x}_j - \overline{X}\right) / (J-1)}{\sum_{j=1}^{J} \left(N_j - 1\right) S_j^2 / (N-1)} \qquad (12)$$

$F(J-1,\ N-J)$ The degree of freedom of a random variable follows the F distribution J-1 and N-1. In the case of a zero predictor, then the p value = 0 for the predictor.

We will now explain step-by-step how to select ANOVA F-test features according to the following algorithm:

Step 1. If you are given standard deviations, proceed to Step 2. Continue to Step 3 if you are given variance to compare.
Step 2. Taking the square root of both standard deviations, we get the variance.
Step 3. To find the *f*-value, divide the largest variance by the smallest variance.
Step 4. Find your degrees of freedom.
Step 5. Check the *f*-value in the table of *f*-values you calculated in Step 3.
Step 6. Compare your calculated value (Step 3) with the table *f*-value in Step 5. The null hypothesis can be rejected if the *F*-table value is smaller than the calculated value.

   Advantage:
(1) *F*-test to decide whether or not the variability between team skill is large than the variability of the observations within the groups.
(2) *F*-test is used both for trying out the hypothesis about the equality of two population variances or the equality of two or greater population means.
(3) In this case, *F*-test is calculated as the ratio between their mean squared errors.
(4) The *F*-tests are surprisingly flexible.
(5) *F*-tests can compare the fits of different models.

   Disadvantage:
(1) ANOVA *F*-test is that it assumes that the organizations have the same, or very similar, standard deviations.
(2) The *F*-test may additionally be tough to determine which team is different.
(3) *F*-test assessments solely capture linear relationships between features and labels.
(4) An extraordinarily correlated feature is given a higher rating, and a lot much less correlated factors are given a decrease score.
(5) The *F*-test using summary data like correlation may additionally be a horrific concept.

3.4.5. Chi-square feature selection

This feature selection approach quantifies the weights of the features making use of the chi-squared statistic and performs the rating of the features. This method can handle both nominal and numerical values. Initially, it conducts the discretization of numerical feature values using various optional discretization methods within the feature processing data. The statistical exams were conducted assuming that the values were usually distributed [47].

Chi-square tests are divided into two types: the goodness-of-fit test and the correlation test. This test determines that the sample data correspond to a population. The second part of the chi-square test of independence examines and compares two variables in a contingency table. It checks whether or not distributions of express variables range from one another. A chi-square test is a statistical test. The formula and circumstances for chi-square tests are given below: [48, 60]

$$x_c^2 = \Sigma \frac{(o_i - E_i)^2}{E_i} \qquad (13)$$

where c is the degree of freedom,
An observed value is O, and
E is an expected value.
Every information object in your facts set will require a calculation if you use the summation symbol.

The steps of the chi-square algorithm are outlined below to explain how it works. [60]

Step 1. A chi-square ($\chi2$) statistic is a measure of the difference between the observed and expected frequencies of the outcomes of a set of events or variables.
Step 2. $\chi2$ depends on the size of the difference between actual and observed values, the degrees of freedom, and the samples size
Step 3. $\chi2$ can be used to test whether two variables are related or independent from one another or to test the goodness-of-fit between an observed distribution and a theoretical distribution of frequencies.

   Advantage:
(1) Chi-square is that it is easier to calculate than some statistics.
(2) Chi-square is used with data that has been measured on a nominal (categorical) scale.
(3) A random variable is being evaluated to determine whether it is independent of another.
(4) The chi-square consists of its robustness with appreciate to distribution of the information.
(5) Its flexibility in handling data from both two groups and multiple.

   Disadvantage:
(1) In the chi-square test, all measured members must be independent, implying that a single entity cannot belong to more than one class.
(2) Chi-square is highly sensitive to sample size. As sample dimension increases, absolute variations emerge as a smaller and smaller share of the predicted value.
(3) A frequency or count of instances needs to be entered rather than a percentage or other type of transformation.
(4) Chi-square requires that the measured elements must be randomly chosen from the entire population.
(5) The data are from a random sample.

3.5. Clustering algorithm

Clustering analysis is an unsupervised machine gaining knowledge of task. The clustering algorithm groups a set of data points into clusters, ensuring that objects within the same cluster share common characteristics. Clustering solely utilizes input data to identify patterns, anomalies, and similarities within the input dataset. Clustering analysis can be divided into a number of types: density-based clustering, connectivity-based clustering, distribution-based clustering, centroids-based clustering, and constraint-based clustering [9].

Machine learning, data mining, pattern recognition, image analysis, and bioinformatics all use clustering to analyze statistical data, among other fields. Clustering of datasets into subsets is a technique for grouping similar objects into exceptional groups. To compare data in each subset based on some distance measure, this is necessary. Below are some clustering algorithms discussed [49, 50, 61].

3.5.1. Hierarchical clustering

Hierarchical clustering includes growing clusters in a predetermined order from top to bottom. Hierarchical clustering is an unattended machine learning algorithm that is used to team the untagged records into a cluster. The algorithm groups similar items into clusters, where each cluster is distinct from others, and the objects within each cluster are mostly similar to one another. The hierarchical clustering creates clusters in a hierarchical tree [62].

There are two types of clustering techniques [49, 50].

➢ **Agglomerative Hierarchical Clustering**
➢ **Divisive Hierarchical Clustering**

• Agglomerative Hierarchical Clustering:

The agglomerative hierarchical clustering follows the bottom-up approach. First, consider each data factor as an individual cluster. Then, iteratively combine pairs of clusters at each step. The following steps provide an understanding of how the agglomerative hierarchical clustering algorithm works. [50].

An observation set of *N* should be grouped as follows:

Step 1. Assign every observation to a single factor cluster, so that if there are N observations, there will be N clusters, each containing simply one observation.
Step 2. Taking the closest clusters (best similar) and combining them into one, we now have N-1 clusters. This can be carried out in a variety of approaches to perceive similar and varied measures.
Step 3. Identify the two closest clusters and merge them into one. There are now N-2 clusters. Linkage techniques using agglomerative clustering can be used for this purpose.
Step 4. The steps 2 and 3 should be repeated until all observations have been clustered into a single group of size N.

• Divisive Hierarchical Clustering:

A division heterarchical clustering differs significantly from an agglomerative tree clustering. When divisive hierarchical clustering is used, each data point in the cluster is separated from the next each time iteration is repeated. There is only one cluster for each disconnected data point. In divisive clustering, all the data are initially grouped into a single cluster, but then each cluster gets split into two children, as a top-down method. Since there are 2N-11 strategies to divide a crew of objects into two groups, it is challenging to compute the best split; subsequently, more than a few heuristics are utilized. The following steps help us understand the workings of the divided hierarchical clustering algorithm [49].

Step 1. Start with one cluster that contains all samples.
Step 2. Each cluster's diameter should be calculated. Samples in a cluster are separated by a diameter, which is the shortest distance between them. The largest diameter cluster C should be chosen for splitting.
Step 3. Determine which sample x from cluster C is the most distinct. Let X leave unique cluster C to construct an independent cluster N (now that cluster C no longer encompasses X). Assign all contributors of cluster C to MC.
Step 4. Step 6 should be repeated until clusters C and N remain the same.
Step 5. Compute the similarity between members of MC and Cluster C and N. Members with the highest similarity in MC should be moved to their similar cluster C and N. Update member C and N.
Step 6. Repeat steps 2, 3, and 4 until the cluster contains the same number of samples. According to the user's specifications.

Advantage:
(1) Easy to put into effect and offers the exceptional result in some cases.
(2) The main edition, the dendrogram, is attractive to users.
(3) It is no longer fundamental to mention the variety of clusters a priori.
(4) They may correspond to meaningful classification.
(5) It is to understand and implement.

Disadvantage:
(1) Does not work very well with missing data.
(2) It does not work well with a variety of data types.
(3) No objective function is directly minimized.
(4) Calculations are expensive and difficult to use with very large datasets.
(5) It is possible to be unaware of a problem when the data have large dimensions.

3.5.2. DBSCAN (Density-based spatial clustering of application with noise clustering)

Solution DBSCAN clustering is an unsupervised learning procedure that identifies characteristic clusters in the data. This approach detects clusters of distinctive sizes and shapes from a large quantity of data. This algorithm involves two parameters: the first one is 'minpts' (minimum number of points), determining the density required for a region to be considered, and the second one utilizes 'minpts' for clustering. Another is eps (ϵ), this parameter works for a distance measure that is used to locate the points [13].

In the following steps, we will implement the density-based spatial clustering algorithm using noise clustering [51, 52, 63].

Step 1. Find all the different sizes and shapes of data neighbor points on the inside eps (ϵ) parameters and identify kernel points with the minpts parameter.

Step 2. A new cluster must be created for each factor point if it is not already assigned.

Step 3. Identify all points with the same density and assign them to the same cluster.

Step 4. The dataset can be iterated over by visiting factors that have not been visited previously. A point that is not associated with a cluster is referred to as a noise point.

Advantage:

(1) While clustering, it is possible to select noise data.
(2) In DBSCAN, clusters can be located of arbitrary size and shape.
(3) A DBSCAN is used to handle outliers within a dataset.
(4) Number of clusters does not require in DBSCAN clustering algorithm.
(5) Clustering algorithm DBSCAN does not require a specific number of clusters.

Disadvantage:

(1) The DBSCAN algorithm expires in clusters with one-of-a-kind densities.
(2) Fails in case of neck type of dataset.
(3) Does not work well in case of high dimensional data.
(4) DBSCAN is its inability to precisely cluster statistics of various density.
(5) In the plot below, we can see two separates clusters of very one-of-a-kind densities.

3.5.3. Shrinkage clustering

Shrinkage clustering is a novel type of clustering algorithm. This algorithm is primarily based on matrix factorization, which allows it to efficiently determine the optimal number of clusters while simultaneously partitioning the data. Cluster evaluation helps pick out exclusive patterns in data and suggests a classification of objects that are functionally comparable or related. The mathematical property of the objective function decreases with each generation in the procedure. With each iteration, the first step is to remove empty clusters. This partial work does not trade input values for the goal function, since objectives characteristic solely as a result of a it is not empty cluster. The second section of this permutation of cluster memberships aims to minimize the objective function. The mathematical condition resulting from the combination of the first and second sections is provided below: [53, 64]

Since

$$||S - AA^T||_{F \geq 0} \tag{14}$$

And

$$||S - AA^T||_F = \sum_{i=1}^{N} \sum_{j \in \{j/A_i = A_j\}}^{N} \left(1 - 2S_{ij}\right) + \sum_{i=1}^{N} \sum_{j=1}^{N} S_{ij}^2 \tag{15}$$

There is a lower bound on the objective function of $-\sum_{i=1}^{N} \sum_{j=1}^{N} S_{ij}^2$. So, a convergence of optimal results is guaranteed, because this method is lower bound algorithm. The shrinkage clustering base algorithms are given below: [64]

Algorithm shrinkage clustering:
Input:$S_{N \times N}$ (Matrix of similarities)
K_0 (Cluster initialization)
Initialization:
a. Generation randomly $A_{N \times K_0}$ (A matrix of cluster assignments)
b. Calculate $\tilde{S} = 1 - 2S$

Continue to repeat

1. Empty clusters should be removed:
   (a) Empty columns should be deleted in A (i.e., $\{ j \mid \sum_{i=1}^{N} A_{ij} = 0 \}$)
2. Permute the cluster membership that minimizes Function 2 the most:
   (a) Calculate $M = \tilde{S}A$
   (b) Calculate v by $v_i = min M_{ij} - \sum_{j=1}^{K} (M \ o \ A)_{ij,}$

where

O represents the element-wise product
(c) Searching the object $\overline{X}$ with the highest optimization potential,

$$\overline{X} = argmin_i v_i$$

(d) Permute the membership of $\overline{X}$ to C' where $C' = argmin_j M_{\overline{X}j}$

Until $\sum_{i=1}^{N} v_i = 0$ or arriving the maximum number of iterations

Output: A (cluster assignment)
Advantage:

(1) The shrinkage clustering discover the cluster memberships.
(2) The main and positive characteristic of Shrinkage clustering is that it cut down the range of clusters when discovering the clustering explanation.
(3) Finding the optimal number of clusters.
(4) When the optimization process is stabilized and the optimal cluster memberships are seen that time clusters automatically decline and become empty.
(5) The number of clusters that do not change in the end is the optimal number of clusters, since it stabilizes the final explanation.

Disadvantage:

(1) The shrinkage clustering algorithms involve computational overhead because the clustering must be performed with different cluster numbers.
(2) Performing shrinkage clustering multiple times helps assess the stability of the explanation.
(3) Implementation burden, since the integration can be laborious if algorithms are programmed.
(4) Shrinkage clustering depended totally on symmetric nonnegative matrix factorization notions.
(5) The clustering requires to be performed with different cluster numbers.

3.5.4. K-means clustering algorithms

K-means clustering is a vector quantization technique that originated in signal processing and goals to subdivide *n* observations within *k* clusters; the cluster model is the one with the nearest mean that accepts every statement [55].

K-means seeks to minimize the sum of squares within clusters by dividing *M* points into *N* dimensions. There is no reason to insist that an explanation have a minimum sum of squares across all partitions, except when *M*, *N* are limited and $K = 2$. We are trying to find as a substitute "local" optimum, explanation so that no motion of a point outside of one cluster to any other will minimize into cluster sum of squares [54].

We are able to understand the following steps of K-means algorithm: [9, 54]

Step 1. First, require to point out the number of clusters, *k*.

Step 2. As a next step, initialize k centroid data points at random and select each data point to cluster.

Step 3. Now calculate the cluster centroids repeatedly.

Step 4. Expectation: set every point too its limited centroid.

Step 5. Maximization: Calculate the new centroid (mean) of every cluster.

Step 6. The centroid positions do not change continuously.

Advantage:

(1) Generalizes to clusters of one-of-a-kind shapes and sizes, such as elliptical clusters.
(2) K-means clustering is convenient to apprehend and implemented.
(3) If there are a big number of variables at that moment k-means work speedily.
(4) Clusters can be changed during recompute of centroids.
(5) Guarantees convergence.

Disadvantage:

(1) Result is powerfully impacted by way of beginning inputs like a variety of clusters (value of k).
(2) K-means has problem clustering statistics where clusters size is change and density.
(3) K-means clustering being dependent on initial values.
(4) Scaling with a number of dimensions.
(5) Predicting the number of clusters is a bit difficult.

3.5.5. Fuzzy C-means clustering

The fuzzy C-means algorithm decides which data points belong to each cluster center based on the range between the cluster center and the data point. In general, the closer the data are to the cluster center, the more likely it is to belong to that precise cluster center. The fuzzy C-mean clusters work to reduce the value of an objective function. The objective function optimizes the weighted within-group sum of square errors (WGSS), as indicated by the following condition: [1, 14, 56]

$$\text{Minimize } J_m(U, W) \equiv \sum_{j=1}^{c} \sum_{i=1}^{N} (\mu_{ij})^m d_{ij}^2 \qquad (16)$$

where N represents the number of patterns in X, C represents the number of clusters, U is the membership function matrix, μ_{ij} is the value of the membership function of the ith pattern involved with j^{th} cluster, d_{ij} is as far as x_i to w_j $d_{ij} = ||x_i - w_j^{(i)}||$, $w_j^{(i)}$ denotes the cluster center of the jth cluster for the ith iteration. w is the cluster center vector, m is the exponent on μ_{ij}.

With the help of the following steps, we can use the fuzzy C-means clustering algorithm [1].

Let $X = \{x_1, x_2, x_3, x_4, \ldots \ldots x_n\}$ be the data points set and $V = \{v_1, v_2, v_3, v_4, \ldots \ldots v_n\}$

Step 1. Randomly selected "c" cluster center.

Step 2. Compute the fuzzy membership μ_{ij}

$$\mu_{ij} = 1 / \sum_{k=1}^{c} \left(d_{ij} / d_{jk} \right)^{\left(\frac{2}{m-1}\right)} \qquad (17)$$

Step 3. Compute the fuzzy center v_j

$$v_n = \left(\sum_{i=1}^{n} (\mu_{ij})^m x_i\right) - \left(\sum_{i=1}^{n} (\mu_{ij})^m\right), \forall_j 1, 2, \ldots \ldots .c \qquad (18)$$

Step 4. Until the minimum is reached, repeat steps 2 and 3 "J" value is achieved or $||U^{(k+1)} - U^k|| < \beta$.

where an iteration step is represented by k,
β is the terminate criterion between [0,1],
$U = (\mu_{ij})_{n*c}$ is the fuzzy membership matrix,
J is the objective function.

Advantage:

(1) A fuzzy C-means clustering technique is a powerful tool for evaluating records and constructing models unsupervised.
(2) The fuzzy C-means algorithm is extensively used.
(3) The FCM makes use of fuzzy separate such that an information factor can be part of to all corporations with separate membership grades between 0 and 1.
(4) A fuzzy clustering method is also more natural than a hard clustering method.
(5) It's not necessary for objects on the borders between the classes to fully belong to one.

Disadvantage:

(1) Fuzzy C-means takes high time for computation.
(2) A measure of how sensitive the initial guess is (speed, local minima).
(3) There is sensitivity to noise and the membership degree of outliers (noisy points) is expected to be low (or even zero).
(4) A priori specification of the number of clusters.
(5) Euclidean distance measures can unequally weight underlying factors.

4. Discussion

In this article, the IoMT is a related environment generated by software program applications, instruments for medical use, and a variety of health systems. It comprises substantial amounts of sensor-based data, encompassing standalone devices and wearables that can be utilized for remote patient monitoring. IoMT is additionally described as healthcare IoMT. Such an ecosystem of IoMT-related devices and sensors have aided to streamline clinical workflow and lead to enhancements in the care of patients. About 60% of worldwide fitness care agencies have already carried out IoMT technology, and some other 27% are anticipated to do so through 2019. Commonly health care is witnessing a paradigm shift like digital conversion puts technologically advanced as nicely as connected products in the consumer's arms additionally gives patients and physicians even in the most remote locations and poorest areas excellent get entry to care of health benefits.

The current state of research in this field is centered around hybrid feature selection methods, specifically focusing on approaches that utilize heuristic algorithms from developmental computation. Examples include the foundation of swarm intelligence-based heuristic systems and various genetic algorithms. In addition, application areas such as image processing, textual content mining, and bioinformatics involve industrial functions with

large-dimensional characteristic spaces. In these contexts, the intelligent hybrid approach structure is of paramount importance for achieving success. Features can have intricate internal shapes or even be unknown in the proceeding. While there is no silver bullet method, filters build on information hypotheses and wrappers build on grasping step-by-step strategies that seem to furnish the first-class outputs. Future research should focus on improving the efficacy and precision of feature subset search methods by integrating and combining previous strategies, encompassing both fine filters and wrappers. Most of the research tends to concentrate on much less extent of datasets on which their approach works. Huge comparative research needs to be pursued in practice to have extra regular outputs.

The most advantageous clustering algorithms are DBSCAN, fuzzy C-means, shrinkage clusters, hierarchical clustering, and k-means. Clustering massive datasets is one of the key applications of k-means in data mining. However, it can only be used for numerical values. The k-mean method could also be improved by including new extensions, thereby enabling us to cluster express records directly using the k-means paradigm without requiring data transformation. Since data mining entails extraordinarily big datasets, scalability is a primary need for data mining methods. Our results of the scan have proven that the k-means methods and SOM applying utilizing of neural network is certainly scalable to extremely massive as well as complex datasets in the shape of both the number of clusters and the range of records. Due to our experiments, the hierarchical clustering algorithm consistently requires fewer reparations than the k-means algorithm, making it faster than the latter. Concerning parallel computing methods to cluster large datasets with hundreds of thousands of objects, we intend to adopt a parallel k-modes approach to distributed databases in our future work. An analysis of such a procedure is useful for a variety of data mining applications, including dividing extraordinarily large sets of heterogeneous objects into smaller groups and controllable homogeneous subsets, which can be analyzed and modeled easily. A broad range of insurance assertions must be examined to identify under-represented concepts, such as fraud.

5. Conclusion

The study aimed to review various biomedical engineering methods and techniques related to the IoMT, feature selection, and clustering algorithms in the context of addressing biological problems during the COVID-19 pandemic. The article first discussed IoMT techniques and evaluated five feature selection techniques: mRMR, random forest, NMIFS, ANOVA *F*-test, and chi-square, highlighting their effectiveness in dealing with different aspects of biological problems during the pandemic. Furthermore, the article reviewed five clustering algorithms: DBSCAN, fuzzy C-means, k-means clustering, shrinkage cluster, and hierarchical clustering, emphasizing their advantages in addressing various factors in bioinformatics engineering problems during the COVID-19 pandemic.

The conclusion drawn from the study suggests that IoMT, feature selection, and clustering techniques are highly valuable and recommended approaches for designing bioinformatics studies related to different biological problems, such as gene detection, dataset analysis, single-cell sequence data imputation, patient treatment and monitoring, and grouping similarity data.

In summary, the article concludes that utilizing IoMT, feature selection, and clustering techniques can effectively address biological problems during the COVID-19 pandemic, providing valuable insights and solutions in the field of bioinformatics. The limitations and comparative studies of the discussed methods were also analyzed to better understand their applicability in fulfilling specific biological objectives under certain restrictions.

6. Future Work

In the future, we will optimize the all-features selection and clustering algorithm. Improve feature selection efficiency and accuracy with advanced algorithms and machine learning techniques. Analyze how different feature selection methods can be effectively combined to address specific challenges in COVID-19 research. Combine multiple clustering algorithms to make your clustering results more accurate and gain deeper insight into complex biomedical datasets using ensemble clustering. On large-scale COVID-19 datasets, evaluate integrated clustering approaches. Investigate how cloud-based IoMT integration can be used with other emerging healthcare technologies, including wearable devices, telemedicine platforms, and electronic medical records. Discover the potential synergies between these technologies to develop comprehensive and personalized solutions for managing and combating COVID-19. In order to ensure seamless collaboration between healthcare systems, interoperability standards and data integration frameworks are crucial.

Recommendations

As biomedical engineers, feature clustering and IoMT provide a number of benefits for COVID-19, including improved data analysis, early detection, personalized treatment, epidemiological modeling, drug discovery, cost efficiency, and interdisciplinary collaboration. Analyzing large amounts of collected data can be accomplished with feature clustering, such as patient information, clinical trials, and epidemiological studies. Patterns and relationships can also be identified using it. COVID-19 infection can be detected and diagnosed early using feature clustering. Clustering of features can assist in detecting and diagnosing COVID-19 infections early. COVID-19 patients can be grouped according to their clinical characteristics and treatment outcomes using feature clustering. With IoMT, treatment response can be monitored in real time, and treatment plans can be adjusted based on the results. COVID-19 can be treated with existing drugs and by using feature clustering to identify possible drug targets. By enabling remote monitoring and personalized treatment, feature clustering and IoMT can reduce healthcare costs. As a result, hospitalizations and costly interventions will be reduced.

Acknowledgement

Ethical Statement

This study does not contain any studies with human or animal subjects performed by any of the authors.

Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

Data Availability Statement

The data that support the findings of this study are openly available in [UCSC Xena] at https://xenabrowser.net/datapages/, in [National Cancer Institute] at https://www.cancer.gov/ccg/access-data, and in [NCBI] at https://www.ncbi.nlm.nih.gov/geo/.

References

[1] Alharbi, E., & Tiun, S. (2015). A hybrid method of linguistic features and clustering approach for identifying biomedical named entities. *Asian Journal of Applied Sciences*, *8*(3), 210–216. https://doi.org/10.3923/ajaps.2015.210.216

[2] Singh, R. P., Javaid, M., Haleem, A., Vaishya, R., & Ali, S. (2020). Internet of Medical Things (IoMT) for orthopaedic in COVID-19 pandemic: Roles, challenges, and applications. *Journal of Clinical Orthopaedics and Trauma*, *11*(4), 713–717. https://doi.org/10.1016/j.jcot.2020.05.011

[3] Schäfer, S. K., Sopp, M. R., Schanz, C. G., Staginnus, M., Göritz, A. S., & Michael, T. (2020). Impact of COVID-19 on public mental health and the buffering effect of a sense of coherence. *Psychotherapy and Psychosomatics*, *89*(6), 386–392. https://doi.org/10.1159/000510752

[4] Vishnu, S., Ramson, S. J., & Jegan, R. (2020). Internet of medical things (IoMT) - An overview. In *2020 5th International Conference on Devices, Circuits and Systems,* 101–104. https://doi.org/10.1109/icdcs48716.2020.243558

[5] De la Torre Díez, I., Alonso, S. G., Hamrioui, S., Cruz, E. M., Nozaleda, L. M., & Franco, M. A. (2018). IoT-based services and applications for mental health in the literature. *Journal of Medical Systems*, *43*, 1–6. https://doi.org/10.1007/s10916-018-1130-3

[6] Jain, A. K., Duin, R. P. W., & Mao, J. (2000). Statistical pattern recognition: A review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *22*(1), 4–37.

[7] Estévez, P. A., Tesmer, M., Perez, C. A., & Zurada, J. M. (2009). Normalized mutual information feature selection. *IEEE Transactions on Neural Networks*, *20*(2), 189–201.

[8] Jović, A., Brkić, K., & Bogunović, N. (2015). A review of feature selection methods with applications. In *38th International Convention on Information and Communication Technology, Electronics and Microelectronics,* 1200–1205. https://doi.org/10.1109/mipro.2015.7160458

[9] Singh, L., Singh, S., & Dubey, P. K. (2010). Applications of clustering algorithms and self-organizing maps as data mining and business intelligence tools on real world data sets. In *International Conference on Methods and Models in Computer Science,* 27–33. https://doi.org/10.1109/icm2cs.2010.5706714

[10] Xu, R., & Wunsch, D. C. (2010). Clustering Algorithms in Biomedical Research: A Review. *IEEE Reviews in Biomedical Engineering*, *3*, 120–154. https://doi.org/10.1109/rbme.2010.2083647

[11] Park, C., Took, C. C., & Seong, J. K. (2018). Machine learning in biomedical engineering. *Biomedical Engineering Letters*, *8,* 1–3.

[12] Keyvanpour, M., & Serpush, F. (2019). ESLMT: A new clustering method for biomedical document retrieval. *Biomedical Engineering/Biomedizinische Technik*, *64*(6), 729–741. https://doi.org/10.1515/bmt-2018-0068

[13] Deng, D. (2020). Application of DBSCAN algorithm in data sampling. *Journal of Physics: Conference Series*, *1617*(1), 012088. https://doi.org/10.1088/1742-6596/1617/1/012088

[14] Li, X., Lu, X., Tian, J., Gao, P., Kong, H., & Xu, G. (2009). Application of fuzzy c-means clustering in data analysis of metabolomics. *Analytical Chemistry*, *81*(11), 4468–4475. https://doi.org/10.1021/ac900353t

[15] Khan, S. S., & Ahmad, A. (2004). Cluster center initialization algorithm for K-means clustering. *Pattern Recognition Letters*, *25*(11), 1293–1302. https://doi.org/10.1016/j.patrec.2004.04.007

[16] Traylor, T., Straub, J., & Snell, N. (2019). Classifying fake news articles using natural language processing to identify in-article attribution as a supervised learning estimator. In *2019 IEEE 13th International Conference on Semantic Computing*, 445–449.

[17] Nasteski, V. (2017). An overview of the supervised machine learning methods. *Horizons. B.*, *4*, 51–62. https://doi.org/10.20544/HORIZONS.B.04.1.17.P05

[18] Mallik, S., Seth, S., Bhadra, T., Tomar, N., & Zhao, Z. (2019). A Multi-classifier Model to Identify Mitochondrial Respiratory Gene Signatures in Human Cancer. In *2019 IEEE International Conference on Bioinformatics and Biomedicine*, 1928–1935. https://doi.org/10.1109/BIBM47256.2019.8982945

[19] Roy, A., Banerjee, S., Bhatt, C., Badr, Y., & Mallik, S. (2018). Hybrid group recommendation using modified termite colony algorithm: A context towards big data. *Journal of Information & Knowledge Management*, *17*(02), 1850019.

[20] Solorio-Fernández, S., Carrasco-Ochoa, J. A., & Martínez-Trinidad, J. F. (2020). A review of unsupervised feature selection methods. *Artificial Intelligence Review*, *53*(2), 907–948. https://doi.org/10.1007/s10462-019-09682-y

[21] Vinh, L. T., Lee, S., Park, Y. T., & d' Auriol, B. J. (2012). A novel feature selection method based on normalized mutual information. *Applied Intelligence*, *37*(1), 100–120. https://doi.org/10.1007/s10489-011-0315-y

[22] Bhalerao, S., & Parab, S. (2010). Choosing statistical test. *International Journal of Ayurveda Research*, *1*(3), 187.

[23] Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., & Smyth, G. K. (2015). Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research*, *43*(7), e47–e47.

[24] Kim, T. K. (2017). Understanding one-way ANOVA using conceptual figures. *Korean Journal of Anesthesiology*, *70*(1), 22–26. https://doi.org/10.4097/kjae.2017.70.1.22

[25] Kim, T. K. (2015). T test as a parametric statistic. *Korean Journal of Anesthesiology*, *68*(6), 540–546. https://doi.org/10.4097/kjae.2015.68.6.540

[26] Liang, G., Fu, W., & Wang, K. (2019). Analysis of t-test misuses and SPSS operations in medical research papers. *Burns & Trauma*, *7*. https://doi.org/10.1186/s41038-019-0170-3

[27] Fraiman, D., & Fraiman, R. (2018). An ANOVA approach for statistical comparisons of brain networks. *Scientific Reports*, *8*(1), 1–14. https://doi.org/10.1038/s41598-018-23152-5

[28] Skaik, Y. A. (2015). The bread and butter of statistical analysis "t-test": Uses and misuses. *Pakistan Journal of Medical Sciences*, *31*(6), 1558–1559. https://doi.org/10.12669/pjms.316.8984

[29] Sánchez-Maroño, N., Alonso-Betanzos, A., & Tombilla-Sanromán, M. (2007). Filter methods for feature selection – A comparative study. In *International Conference on Intelligent Data Engineering and Automated Learning, 4881*, 178–187. https://doi.org/10.1007/978-3-540-77226-2_19

[30] Siddiqi, M. A., & Pak, W. (2020). Optimizing Filter-Based Feature Selection Method Flow for Intrusion Detection System. *Electronics*, *9*(12), 2114. https://doi.org/10.3390/electronics9122114

[31] Adler, J., & Parmryd, I. (2010). Quantifying colocalization by correlation: The Pearson correlation coefficient is superior to the Mander's overlap coefficient. *Cytometry Part A*, *77A*(8), 733–742. https://doi.org/10.1002/cyto.a.20896

[32] Maldonado, S., & Weber, R. (2009). A wrapper method for feature selection using Support Vector Machines. *Information Sciences*, *179*(13), 2208–2217. https://doi.org/10.1016/j.ins.2009.02.014

[33] Chen, R. C., Dewi, C., Huang, S. W., & Caraka, R. E. (2020) Selecting critical features for data classification based on machine learning methods. *Journal of Big Data*, *7*(1), 1–26. https://doi.org/10.1186/s40537-020-00327-4

[34] Sahoo, K. K., Ghosh, R., Mallik, S., Roy, A., Singh, P. K., & Zhao, Z. (2023). Wrapper-based deep feature optimization for activity recognition in the wearable sensor networks of healthcare systems. *Scientific Reports*, *13*(1), 965. https://doi.org/10.1038/s41598-022-27192-w

[35] Sen, S., Saha, S., Chatterjee, S., Mirjalili, S., & Sarkar, R. (2021). A bi-stage feature selection approach for COVID-19 prediction using chest CT images. *Applied Intelligence*, *51*, 8985–9000.

[36] Afzal, A., Ansari, Z., Alshahrani, S., Raj, A. K., Kuruniyan, M. S., Saleel, C. A., & Nisar, K. S. (2021). Clustering of COVID-19 data for knowledge discovery using c-means and fuzzy c-means. *Results in Physics*, *29*.

[37] Billah, M., & Waheed, S. (2020). Minimum redundancy maximum relevance (mRMR) based feature selection from endoscopic images for automatic gastrointestinal polyp detection. *Multimedia Tools and Applications*, *79*, 23633–23643. https://doi.org/10.1007/s11042-020-09151-7

[38] Berrendero, J. R., Cuevas, A., & Torrecilla, J. L. (2016). The mRMR variable selection method: A comparative study for functional data. *Journal of Statistical Computation and Simulation*, *86*(5), 891–907.

[39] Cai, Y., Huang, T., Hu, L., Shi, X., Xie, L., & Li, Y. (2012). Prediction of lysine ubiquitination with mRMR feature selection and analysis. *Amino Acids*, *42*(4), 1387–1395. https://doi.org/10.1007/s00726-011-0835-0

[40] Ding, C., & Peng, H. (2005). Minimum redundancy feature selection from microarray gene expression data. *Journal of Bioinformatics and Computational Biology*, *3*(02), 185–205. https://doi.org/10.1109/ICOSC.2019.8665593

[41] Mehmood, R., Shahzad, W., & Ahmed, E. (2017) Maximum relevancy minimum redundancy-based feature subset selection using ant colony optimization. *Journal of Applied Environmental Biological Sciences*, *7*(4), 118–130.

[42] Peng, H., Long, F., Ding, C. (2005). Feature selection based on mutual information: Criteria of max-dependency, maxrelevance and min-redundancy. *IEEE Transactionsons on Pattern Analysis and Machine Intelligence*, *27*(8), 1226–1238.

[43] Kumar, S. S., & Shaikh, T. (2017). Empirical evaluation of the performance of feature selection approaches on random forest. In *2017 International Conference on Computer and Applications,* 227–231. https://doi.org/10.1109/COMAPP.2017.8079769

[44] Kursa, M. B., & Rudnicki, W. R. (2011). The all relevant feature selection using random forest. *arXiv Preprint: 1106.5112*.

[45] Sarica, A., Cerasa, A., & Quattrone, A. (2017). Random forest algorithm for the classification of neuroimaging data in Alzheimer's Disease: A systematic review. *Frontiers in Aging Neuroscience*, *9*, 329. https://doi.org/10.3389/fnagi.2017.00329

[46] Dhanya, R., Paul, I. R., Akula, S. S., Sivakumar, M., & Nair, J. J. (2020). F-test feature selection in Stackin ensemble model for breast cancer prediction. *Procedia Computer Science*, *171*, 1561–1570. https://doi.org/10.1016/j.procs.2020.04.167

[47] Bahassine, S., Madani, A., Al-Sarem, M., & Kissi, M. (2018). Feature selection using an improved Chi-square for Arabic text classification. *Journal of King Saud University – Computer and Information Sciences*, *32*(2), 225–231. https://doi.org/10.1016/j.jksuci.2018.05.010

[48] Jin, C., Ma, T., Hou, R., Tang, M., Tian, Y., Al-Dhelaan, A., & Al-Rodhaan, M. (2015). Chi-square statistics feature selection based on term frequency and distribution for text categorization. *IETE Journal of Research*, *61*(4), 351–362. https://doi.org/10.1080/03772063.2015.1021385

[49] Bandyopadhyay, S. & Coyle, E. J. (2003). An energy efficient hierarchical clustering algorithm for wireless sensor networks. In *IEEE INFOCOM 2003. Twenty-Second Annual Joint Conference of the IEEE Computer and Communications Societies*, *3,* 1713–1723. https://doi.org/10.1109/infcom.2003.1209194

[50] Nazari, Z., Kang, D., Asharif, M. R., Sung, Y., & Ogawa, S. (2015). A new hierarchical clustering algorithm. In *International Conference on Intelligent Informatics and Biomedical Sciences,* 148–152. https://doi.org/10.1109/iciibms.2015.7439517

[51] Birant, D., & Kut, A. (2007). ST-DBSCAN: An algorithm for clustering spatial – temporal data. *Data & Knowledge Engineering*, *60*(1), 208–221. https://doi.org/10.1016/j.datak.2006.01.013

[52] Tran, T. N., Drab, K., & Daszykowski, M. (2013). Revised DBSCAN algorithm to cluster data with dense adjacent clusters. *Chemometrics and Intelligent Laboratory Systems*, *120*, 92–96. https://doi.org/10.1016/j.chemolab.2012.11.006

[53] Ishwaran, H., & Sunil Rao, J. S. (2008). Clustering gene expression profile data by selective shrinkage. *Statistics & Probability Letters*, *78*(12), 1490–1497. https://doi.org/10.1016/j.spl.2008.01.003

[54] Ghoshal, S., Nohria, N., & Wong, M. A. (1979). A K-means clustering algorithm. Algorithm AS 136. *Applied Statistics*, *28*, 100–108. https://doi.org/10.2307/2346830

[55] Likas, A., Vlassis, N., & Verbeek, J. (2003). The global k-means clustering algorithm. *Pattern Recognition*, *36*(2), 451–461. https://doi.org/10.1016/s0031-3203(02)00060

[56] Bezdek, J. C., Ehrlich, R., & Full, W. (1984). FCM: The fuzzy c-means clustering algorithm. *Computers and Geosciences*, *10*, 191–203. https://doi.org/10.1016/0098-3004(84)90020-7

[57] Wang, Z., Xiao, X., & Rajasekaran, S. (2020). Novel and efficient randomized algorithms for feature selection. *Big Data Mining and Analytics*, *3*(3), 208–224.

[58] Bugata, P., & Drotar, P. (2020). On some aspects of minimum redundancy maximum relevance feature selection. *Science China Information Sciences*, *63*, 1–15. https://doi.org/10.1007/s11432-019-2633-y

[59] Enireddy, V., PhaniKumar, D. V. V. S., & Kishore, G. (2016). Application of fisher score and mRMR techniques for feature selection in compressed medical images. *International Journal of Engineering and Technology*, *7*(6), 2109–2121.

[60] Elssied, N. O. F., Ibrahim, O., & Osman, A. H. (2014). A novel feature selection based on one-way Anova F-test for e-mail spam classification. *Research Journal of Applied Sciences, Engineering and Technology*, *7*(3), 625–638.

[61] Rodriguez, M. Z., Comin, C. H., Casanova, D., Bruno, O. M., Amancio, D. R., Costa, L. D. F., & Rodrigues, F. A. (2019). Clustering algorithms: A comparative approach. *PloS One*, *14*(1), e0210236. https://doi.org/10.1371/journal.pone.0210236

[62] Sousa, L., & Gama, J. (2014). The application of hierarchical clustering algorithms for recognition using biometrics of the hand. *International Journal of Advanced Engineering Research and Science*, *1*(7), 14–24.

[63] Kumar, K. M., & Reddy, A. R. M. (2016). A fast DBSCAN clustering algorithm by accelerating neighbor searching using Groups method. *Pattern Recognition*, *58*, 39–48. https://doi.org/10.1016/j.patcog.2016.03.008

[64] Hu, C. W., Li, H., & Qutub, A. A. (2018). Shrinkage Clustering: A fast and size-constrained clustering algorithm for biomedical applications. *BMC Bioinformatics*, *19*, 1–9. https://doi.org/10.1186/s12859-018-2022-8