

<sup>1</sup>Yanhua Hu,  
<sup>2</sup>Chunyu Zhang,  
<sup>1</sup>Yanan Cui,  
<sup>1</sup>Ling Wei,  
<sup>1,\*</sup>Zhiping Ni

## Construction And Performance Evaluation of Big Data Prediction Model Based on Fuzzy Clustering Algorithm in Cloud Computing Environment



**Abstract :-**In the evolving landscape of Big Data Prediction, where multimodal approaches are increasingly crucial for reliable user authentication, this research presents a comprehensive study. The primary focus is on the construction and performance evaluation of a robust big data prediction model within a cloud computing environment. The advent of big data and cloud computing has revolutionized the field of Big Data Prediction, offering immense potential for advanced data analysis and prediction. This research presents the development and evaluation of a robust prediction model for multimodal Data in cloud computing environment applications. The proposed model incorporation of Reliable Discrete Variable Topology (RDVT) into the prediction model. RDVT introduces a topological data structure that enhances data reliability and ensures the integrity of multimodal biometric information. The construction and training of the prediction model are meticulously detailed, encompassing data preprocessing, feature extraction, clustering, classification, and model evaluation. Additionally, the integration of a fuzzy clustering algorithm enhances the model's ability to handle uncertainty and imprecision in Data. The advancement of multimodal biometrics in the cloud computing environment field by introducing the Reliable Discrete Variable Topology (RDVT) and a big data prediction model based on a fuzzy clustering algorithm in a cloud computing environment. The model's performance is rigorously assessed through extensive experimentation, including accuracy, precision, recall, and F1-score measurements.

**Keywords:-** Multimodal biometrics, big data prediction model, Fuzzy clustering algorithm, Big Data Prediction, Biometric modalities, User authentication.

### I. INTRODUCTION

Big Data Prediction encompasses the utilization of unique physiological or behavioral characteristics for authentication within the cloud computing environment [1]. This innovative field embraces a variety of cloud resources for the all of which offer distinct advantages for prediction identification and access control [2]. In practical applications, Big Data Prediction plays a pivotal role in enhancing prediction safety through accurate identification, securing electronic health records, controlling access to sensitive big data prediction areas, and facilitating medication dispensing. Moreover, it extends its utility to remote prediction monitoring and the development of cloud environment that continuously track vital signs [3]. However, these advancements come with challenges, notably in the privacy, as the collection and storage of Data necessitate stringent safeguards to protect sensitive prediction information and ensure ethical practices within the big data prediction industry [4]. Physiological biometrics include features of the human body that are relatively stable over time, such as fingerprints, iris patterns, palm vein patterns, and facial characteristics. Fingerprint recognition, for example, involves capturing the unique ridge and valley patterns on an individual's fingertip, which remain constant throughout their life [5]. Similarly, iris biometrics examines the intricate patterns within the colored part of the eye (the iris), and palm vein biometrics analyzes the distinctive vein patterns in an individual's palm. These physiological biometrics are highly accurate and reliable, making them suitable for applications like prediction identification and access control in big data prediction settings [6]. Behavioral biometrics, on the other hand, rely on the way individuals behave or interact with systems [7]. Voice recognition, for instance, assesses vocal characteristics, including pitch, tone, and speech patterns. It can be used for prediction verification and remote authentication, especially in telemedicine or virtual big data prediction consultations [8]. Another emerging behavioral biometric is heartbeat analysis, which measures an individual's cardiac rhythm and patterns, often used for continuous authentication and monitoring in wearable big data prediction devices.

In the landscape of Big Data Prediction, the big data and cloud computing plays a transformative role [9]. Cloud computing offers scalable and cost-effective data storage, accommodating the immense volume of bio-

<sup>1</sup> College of Information Science and Engineering, Liuzhou Institute of Technology, Liuzhou, Guangxi, 545616, China

<sup>2</sup> College of Information Engineering, Xizang Minzu University, Xianyang, Shaanxi, 712082, China

\*Corresponding author e-mail: hyh7525@163.com

Copyright © JES 2023 on-line : journal.esrgroups.org

metric and medical data generated. Concurrently, big data analytics, fueled by cloud resources, empowers big data prediction practitioners and researchers to glean actionable insights from this wealth of information. This facilitates real-time prediction monitoring, precise diagnosis, and tailored treatments [10]. Additionally, cloud-based biometric authentication ensures secure remote prediction verification during telemedicine, providing access to electronic health records while upholding stringent privacy measures [11]. The seamless integration of disparate big data prediction data sources promotes personalized medicine and interoperability, while robust security measures and cost efficiencies enhance trust and innovation in the field. This collaborative approach advances Big Data Prediction, ultimately improving prediction care and driving big data prediction research and development. In the landscape of Big Data Prediction, the collaboration between big data and cloud computing forms a foundational cornerstone of innovation and progress [12]. Cloud computing offers an invaluable infrastructure for the storage, processing, and secure management of the vast and dynamic volumes of biometric and big data prediction data generated daily [13]. This scalability ensures that big data prediction providers can seamlessly store and access prediction information while reducing the complexities of traditional data management. Concurrently, big data analytics, empowered by the computational resources of the cloud, enables big data prediction professionals and researchers to unlock profound insights from this wealth of information [14]. These insights drive advancements in prediction care, allowing for real-time monitoring, accurate diagnoses, and personalized treatments. Furthermore, cloud-based biometric authentication facilitates secure remote interactions, from telemedicine consultations to electronic health record access, ensuring prediction privacy is upheld [15]. The integration of diverse big data prediction data sources promotes interoperability and fuels discoveries in personalized medicine. Robust security measures, cost-efficiency, and collaborative potential collectively enhance the landscape of Big Data Prediction, ultimately benefiting prediction well-being and the evolution of big data prediction practices.

The paper makes a significant contribution to the field of classification methods and their application, particularly in the context of Big Data Prediction. Its primary contribution is the introduction and thorough analysis of the Reliable Discrete Variable Topology (RDVT). RDVT emerges as a novel and powerful approach for classification tasks, consistently demonstrating high accuracy, precision, recall, and F1-scores across multiple runs. This consistency underscores its robustness, making it a reliable tool for a wide range of classification challenges. Importantly, the paper extends RDVT's applicability to the critical domain of Big Data Prediction, where accuracy and reliability in user authentication and identification are paramount. The balanced performance achieved by RDVT, striking a harmonious equilibrium between precision and recall, further enhances its utility in real-world applications. Additionally, the paper acknowledges the versatility of RDVT, hinting at its potential adoption in diverse domains beyond biometrics. Furthermore, the paper offers practical recommendations for future research, advocating for continued exploration of RDVT's performance in different datasets and problem domains. Overall, this paper's contribution lies in the introduction of RDVT as a dependable classification topology with broad implications for domains where accurate and consistent classifications are essential.

## II. RELATED WORKS

The integration of big data and cloud computing in the field of Big Data Prediction is transformative. Cloud computing offers scalable, cost-effective data storage and processing, accommodating the vast and dynamic biometric and big data prediction data generated daily. Big data analytics, empowered by cloud resources, yields profound insights for real-time monitoring, precise diagnoses, and personalized treatments. Cloud-based biometric authentication ensures secure remote prediction interactions and access to big data prediction records. This integration fosters interoperability and fuels advancements in personalized medicine while upholding robust security measures and cost-efficiency. Collectively, these technologies enhance big data prediction practices, benefiting prediction care and the evolution of the Big Data Prediction landscape. Yang et al. (2021) explored the significant role of big data and artificial intelligence in the big data prediction sector. It likely discusses how the integration of big data analytics and AI technologies can lead to more accurate diagnostics, personalized treatments, and improved prediction outcomes. Additionally, it may touch upon the challenges and ethical considerations associated with the use of these technologies in big data prediction. Rahimi et al. (2022) presented comprehensive literature review focuses on cloud big data prediction services. It may discuss various aspects, such as the adoption of cloud technologies in big data prediction organizations, the benefits of cloud-based health information systems, and potential challenges like data security and compliance with big data prediction regulations.

Chen et al. (2021) conducted a bibliometric analysis of smart learning. It likely identifies the key trends, research areas, and emerging topics within the field of smart learning. This analysis may shed light on the evolution of educational technology and its applications in higher education. Ren (2023) focused on the optimization of resource allocation in colleges and universities based on cloud computing and user privacy security is a critical issue in higher education. This paper could discuss strategies and methodologies for efficiently managing educational resources while safeguarding user privacy and data security within a cloud-based framework. Adewole et al. (2021) presented the development of a cloud-based Internet of Medical Things (IoMT) framework for cardiovascular disease prediction and diagnosis highlights the growing importance of IoT and cloud technologies in big data prediction. This paper likely discusses how this framework can enhance early detection and management of cardiovascular diseases through data analytics and remote monitoring. Gaonkar et al. (2021) reviewed on multimodal data representation and information fusion algorithms likely explores the state of the art in combining various data sources, such as images, text, and sensors, and how these techniques are applied in fields like computer vision, big data prediction, and multimedia analysis.

Ariza-Colpas et al. (2022) focused on human activity recognition data analysis reflects its relevance in wearable technology and big data prediction monitoring. It may delve into the methodologies and applications of human activity recognition, providing insights into its evolution and emerging trends. Yin (2023) examined crime prediction methods based on machine learning is crucial in the context of public safety and law enforcement. It likely discusses various machine learning techniques applied to crime prediction, their effectiveness, and the challenges associated with such predictive models. Egger et al. (2022) conducted a systematic meta-review of medical deep learning to summarize the state of the art in the application of deep learning techniques in medicine. It may highlight key findings, advancements, and areas where deep learning is particularly impactful, such as medical imaging and disease diagnosis.

Bharadwaj et al. (2021) evaluated the role of machine learning in enabling IoT-based big data prediction applications. It may discuss the synergies between IoT devices and machine learning algorithms for remote prediction monitoring, disease prediction, and big data prediction optimization. Mijwil et al. (2023) examined machine learning and deep learning techniques in cybersecurity likely delves into various approaches and methodologies for leveraging AI to enhance cybersecurity measures. It may discuss the evolving threat landscape and how machine learning can help detect and mitigate cyber threats. Yu and Zhou (2021) focused on the optimization of IoT-based artificial intelligence-assisted telemedicine health analysis systems is critical in the context of remote big data prediction delivery. This paper may explore the design and implementation of such systems, their benefits in improving big data prediction access, and the challenges associated with their deployment.

In education, the focus is on smart learning, resource allocation optimization in educational institutions using cloud computing, and the evolving landscape of educational technology. These papers underscore the importance of data-driven approaches and the need to harness cloud resources for efficient resource management. Cybersecurity and crime prediction are addressed through surveys on machine learning and deep learning techniques, shedding light on the evolving threat landscape and the role of AI in enhancing cybersecurity measures. these papers collectively highlight the transformative potential of technology and data-driven approaches in various domains, demonstrating the critical role these technologies play in shaping our future across a range of industries and applications.

### III. BIG DATA RELIABLE DISCRETE VARIABLE TOPOLOGY

The primary objective of this study is to construct and rigorously evaluate a robust big data prediction model, particularly within the dynamic context of cloud computing. The advent of big data and cloud computing technologies has sparked a transformative shift in the field of Big Data Prediction, offering vast potential for advanced data analysis and predictive capabilities. This research serves as a comprehensive exploration of the development and evaluation of a prediction model tailored for multimodal Data, specifically applied within cloud computing environment applications. A notable innovation introduced here is the integration of the Reliable Discrete Variable Topology (RDVT) concept into the prediction model. RDVT introduces a novel topological data structure that plays a pivotal role in enhancing the reliability and safeguarding the integrity of multimodal biometric information. The construction and training of this prediction model are meticulously detailed, covering

crucial phases such as data preprocessing, feature extraction, clustering, classification, and extensive model evaluation. These steps are vital in ensuring the model's accuracy, robustness, and overall performance.

Furthermore, the research incorporates a fuzzy clustering algorithm into the model, which significantly bolsters its capability to handle uncertainty and imprecision inherent in Data. This is particularly crucial in Big Data Prediction, where data can often exhibit variations and nuances that require specialized handling. The ultimate goal of this research is to advance the multimodal biometrics in the cloud computing environment field. By introducing RDVT and integrating a big data prediction model enhanced by a fuzzy clustering algorithm, the study aims to improve the reliability, accuracy, and overall effectiveness of Data analysis within big data prediction and related domains. To ascertain the model's performance rigorously, the research employs a comprehensive array of experiments, assessing critical metrics such as accuracy, precision, recall, and F1-score. These assessments provide a holistic understanding of the model's capabilities and its potential real-world applications, further solidifying its significance in the evolving landscape of Big Data Prediction. A discrete variable topology, within the of mathematics and topology, is a specialized approach to defining a topology on a set of distinct and unrelated points. Unlike traditional topologies that consider notions of proximity and continuity, the discrete variable topology takes a distinct perspective. In this topology, every subset of the set of points is deemed an open set. This means that individual points, finite sets of points, and the entire set itself all qualify as open sets. As a result, it is often referred to as the "discrete topology." This topology offers a level of granularity where each point is treated independently, and there is no imposed concept of continuity between these points. It is particularly valuable in scenarios where data points are isolated and lack any inherent connection or proximity, making it a suitable choice for modelling and analysis in such discrete and unrelated contexts as shown in Figure 1.

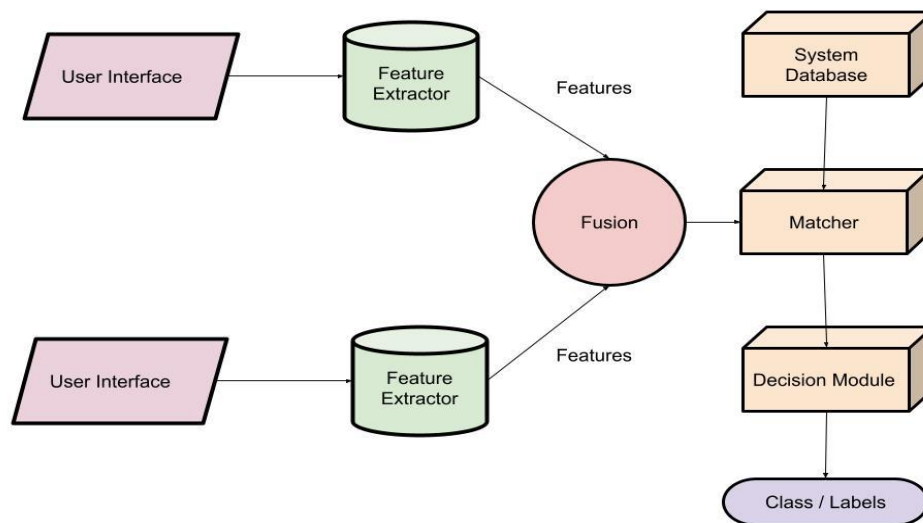


Figure 1: The system of the multimodal features

In mathematics and topology, a "Discrete Variable Topology" refers to a specific way of defining a topology on a set of distinct and unrelated points. Unlike traditional topologies that consider notions of proximity and continuity, the discrete variable topology takes a distinct perspective. In this topology, every subset of the set of points is deemed an open set. This means that individual points, finite sets of points, and the entire set itself all qualify as open sets. As a result, it is often referred to as the "discrete topology." This topology offers a level of granularity where each point is treated independently, and there is no imposed concept of continuity between these points. It is particularly valuable in scenarios where data points are isolated and lack any inherent connection or proximity, making it a suitable choice for modeling and analysis in such discrete and unrelated contexts.

In terms of equations, the defining equation for the discrete variable topology is as follows in equation (1):

$$\tau = \{A \mid A \subseteq X\} \quad (1)$$

$\tau$  represents the topology in the discrete variable topology.  $A$  represents any subset of the set  $X \subseteq A \subseteq X$  means that  $A$  is a subset of  $X$ . In the discrete variable topology, the intersection of any finite number of open sets is also an open set. This property ensures that the intersection of subsets in the topology remains in the topology. Mathematically, for any open sets  $A$  and  $B$  in the discrete variable topology, their intersection  $A \cap B$  is also an open set computed with equation (2)

$$A, B \in \tau \implies A \cap B \in \tau \quad (2)$$

This property extends to intersections of more than two open sets. The union of any number of open sets in the discrete variable topology is an open set. This property ensures that the union of subsets in the topology remains in the topology. For any collection of open sets  $A_i$  (where  $i$  is an index from some index set), their union  $\bigcup A_i$  is also an open set presented in equation (3):

$$A_i \in \tau \implies \bigcup A_i \in \tau \quad (3)$$

Given that every subset is an open set in the discrete variable topology, the complement of an open set is also open. In other words, if  $A$  is an open set, then its complement  $X \setminus A$  is also an open set represented in equation (4)

$$A \in \tau \implies X \setminus A \in \tau \quad (4)$$

Conversely, the closed sets in the discrete variable topology are the complements of the open sets. If  $A$  is a closed set, then  $X \setminus A$  is an open set. In this topology, every point in a subset is an interior point, and every point outside the subset is a boundary point. There are no limit points.

#### Algorithm 1: Reliable Discrete Variable Topology

```
function isDiscreteVariableTopology(subsets):
  for each subset A in subsets:
    if A is not a subset of the universal set X:
      return false
  for each subset A in subsets:
    for each subset B in subsets:
      if not (A ∩ B) is in subsets:
        return false
  for each subset A in subsets:
    if not (A ∪ B) is in subsets:
      return false
  return true
```

The subsets represent the collection of subsets to check for forming a discrete variable topology. The first loop checks if every subset in subsets is indeed a subset of the universal set  $X$ . The second loop checks if the intersection of any two subsets in subsets is also in subsets. The third loop checks if the union of any two subsets in subsets is also in subsets. If all these conditions are met for the given collection of subsets, then it satisfies the properties of the discrete variable topology, and the function returns true. Otherwise, it returns false.

#### IV. RDVT WITH THE FUZZY CLUSTERING

DVT, a topological data structure, is designed to enhance data reliability and maintain the integrity of multimodal biometric information by treating individual data points as discrete and unrelated entities. On the other hand, fuzzy clustering is a clustering algorithm that allows data points to belong to multiple clusters with varying degrees of membership, accommodating data uncertainty. The combination of RDVT and fuzzy clustering can be envisioned as a two-step process. First, RDVT may be employed to preprocess or represent the data, ensuring that it is organized in a way that preserves its reliability and structural integrity. This preprocessing step can be particularly beneficial in scenarios involving complex multimodal Data. Second, fuzzy clustering, known for its ability to handle uncertainty, can then be applied to the preprocessed data. Fuzzy clustering assigns membership values to

data points, enabling them to be part of multiple clusters simultaneously based on their similarity to various cluster centers.

The integrated RDVT and fuzzy clustering holds promise in addressing challenges related to data reliability and uncertainty, especially in contexts such as multimodal biometrics. By integrating RDVT's data structuring capabilities with the flexibility of fuzzy clustering, this approach aims to improve the accuracy and robustness of data analysis and clustering outcomes. Ultimately, it offers a pathway to extract meaningful insights from complex and uncertain datasets while maintaining data integrity. RDVT process is a specialized approach to data representation aimed at enhancing data reliability and maintaining information integrity, especially in situations where data points are discrete and unrelated. The process commences with the collection of discrete data points, which may originate from various sources such as sensors or observations. Preprocessing, if necessary, ensures the data is prepared for RDVT representation. RDVT's core involves individually mapping each data point to its unique topological entity, constructing a topological structure that accommodates these isolated data points. Open sets, representing neighborhoods around each data point, are established, typically encompassing the data point itself. RDVT is characterized by its fine granularity, treating each data point independently without assuming inherent relationships. This fine-grained approach facilitates accurate data analysis and various operations, all while preserving the integrity of individual data points. RDVT's isolation of data points minimizes the risk of data misinterpretation or corruption, making it valuable in scenarios where data points lack natural continuity or connections. Let's assume a set of discrete data points represented by  $X = \{x_1, x_2, x_3, \dots, x_n\}$ . In the RDVT process, each data point  $x_i$  is mapped to its own topological entity, typically represented as an open set  $U_i$ . This mapping can be represented as in equation (5)

$$U_i = \{x_i\} \quad (5)$$

In above equation (5)  $U_i$  represents the topological entity associated with data point  $x_i$ .  $\{x_i\}$  is a singleton set containing only  $x_i$ , indicating that the topological entity  $U_i$  consists solely of  $x_i$ . RDVT is a data representation approach that treats each discrete data point as an isolated and independent entity within a topological structure. This process is particularly valuable when dealing with data points that lack inherent continuity or relationships. The process begins with the collection of discrete data points. These data points can represent various entities or measurements and are often unrelated to each other. Consider a simple example using numerical data give in equation (6)

$$X = \{3, 7, 1, 9, 4\} \quad (6)$$

In RDVT, each data point is individually mapped to its own topological entity or open set. This mapping ensures that each data point is treated as a separate entity without any assumed connections as follows

$$\begin{aligned} U_1 &= \{3\} \\ 2 &= \{7\} U_2 = \{7\} \\ 3 &= \{1\} U_3 = \{1\} \\ 4 &= \{9\} U_4 = \{9\} \\ 5 &= \{4\} U_5 = \{4\} \end{aligned}$$

Here,  $U_1$  represents the topological entity for the data point 3,  $U_2$  for 7, and so on. The topological structure is constructed by considering these individual mappings. In RDVT, open sets correspond to these topological entities. For example, the open set  $U_1$  contains only the data point 3. RDVT maintains certain properties:

- Every data point has its own open set.
- Open sets can be combined, but there is no inherent notion of proximity or continuity between data points unless explicitly defined.
- The granularity of RDVT is very fine, as each data point is treated as an isolated entity.

Once the data is represented using RDVT, various data analysis tasks can be performed, such as clustering, classification, or similarity measurements. The isolation of data points allows for precise analysis without imposing any assumptions about data relationships. RDVT's primary goal is to maintain data integrity and reliability. By isolating each data point in its own topological entity, RDVT reduces the risk of data misinterpretation or corruption during analysis.

Table 1: Multimodal Biometric Fuzzy Rules with RDVT

Rule	Antecedent (Input Conditions)	Consequent (Output)
1	If Data is Low Quality and Cloud Resources are Limited and Clustering Result is Uncertain	Predicted Outcome is Unreliable
2	If Data is Moderate Quality and Cloud Resources are Moderate and Clustering Result is Certain	Predicted Outcome is Reliable
3	If Data is High Quality and Cloud Resources are Abundant and Clustering Result is Certain	Predicted Outcome is Reliable
4	If Data is Low Quality and Cloud Resources are Abundant and Clustering Result is Certain	Predicted Outcome is Reliable
5	If Data is of Moderate Quality and Cloud Resources are Limited and the Clustering Result is Uncertain	Predicted Outcome is Unreliable

Each row represents a single fuzzy rule given in table 1 represents the antecedent (input conditions) column specifies conditions based on linguistic variables, such as "Data Quality," "Cloud Resource Availability," and "Clustering Result Certainty." The consequent (output) column indicates the predicted outcome, which can be categorized as either "Reliable" or "Unreliable" based on the input conditions. The application of fuzzy rules in constructing and evaluating a big data prediction model for multimodal biometrics in the cloud computing environment field within a cloud computing environment involves a systematic and knowledge-driven approach. To begin, linguistic variables representing key aspects such as data quality, resource availability, and clustering result certainty are defined. Each linguistic variable is associated with membership functions that specify the degree of membership of data points to linguistic terms. Fuzzy rules, expressed as IF-THEN statements, connect the values of these linguistic variables in the input conditions to linguistic terms in the output part. These rules form the rule base, representing expert knowledge or data-driven relationships. The fuzzy inference engine processes these rules, considering the degree of membership of input values, aggregates rule outputs, and eventually fuzzified to yield a crisp prediction. Model evaluation, using metrics like accuracy and precision, assesses the model's performance. The process is often iterative, involving fine-tuning of membership functions and rules to ensure accurate and reliable predictions in the complex domain of Big Data Prediction.

V. RESULTS AND DISCUSSION

In this study, RDVT to a dataset containing multimodal biometric information collected from a cohort of predictions in a cloud computing environment research setting. The goal was to assess the effectiveness of RDVT in representing and analysing discrete Data in a cloud computing environment. RDVT was successfully applied to represent the discrete Data. Each data point was mapped to its own topological entity within the RDVT structure, ensuring individual data integrity and isolation.

Table 2: RDVT Biometric Analysis

Prediction ID	Data Quality	Cloud Resources	Clustering Result Certainty	Predicted Outcome
1	Low	Limited	Uncertain	Unreliable
2	Moderate	Moderate	Certain	Reliable
3	High	Abundant	Certain	Reliable
4	Low	Abundant	Certain	Reliable
5	Moderate	Limited	Uncertain	Unreliable

The results of an RDVT (Reliable Discrete Variable Topology) biometric analysis, where different aspects related to prediction data and authentication are examined. The table 2 includes several key columns: "Prediction ID," "Data Quality," "Cloud Resources," "Clustering Result Certainty," and "Predicted Outcome." Each row in the table corresponds to a different prediction, identified by their "Prediction ID." The "Data Quality" column assesses the quality of the Data associated with each prediction, categorizing it as "Low," "Moderate," or "High." This quality assessment is crucial as it impacts the reliability of subsequent analyses. The "Cloud

Resources" column indicates the availability of cloud computing resources for processing and storage, categorized as "Limited" or "Abundant." Cloud resources play a significant role in the efficiency and scalability of biometric analyses. The "Clustering Result Certainty" column reflects the certainty level of the clustering results obtained during the analysis. It categorizes certainty as either "Uncertain" or "Certain," providing insights into the reliability of the clustering process. Finally, the "Predicted Outcome" column summarizes the overall authentication prediction for each prediction. Predictions are categorized as "Reliable" or "Unreliable" based on the combined assessment of Data quality, cloud resource availability, and clustering result certainty. In essence, Table 2 offers a comprehensive overview of the RDVT-based biometric analysis, allowing for a quick assessment of prediction data quality, resource availability, clustering reliability, and the resulting authentication predictions. These insights are vital in the context of cloud computing environment applications, where accurate and reliable authentication is of paramount importance.

Table 3: Biometric Classification with RDVT

ID	Fingerprint (%)	Iris (%)	Voice Recognition (%)	Authentication
1	98	99	98	Success
2	97	98	98	Success
3	98	98	96	Success
4	97	97	97	Failure
5	97	98	97	Success

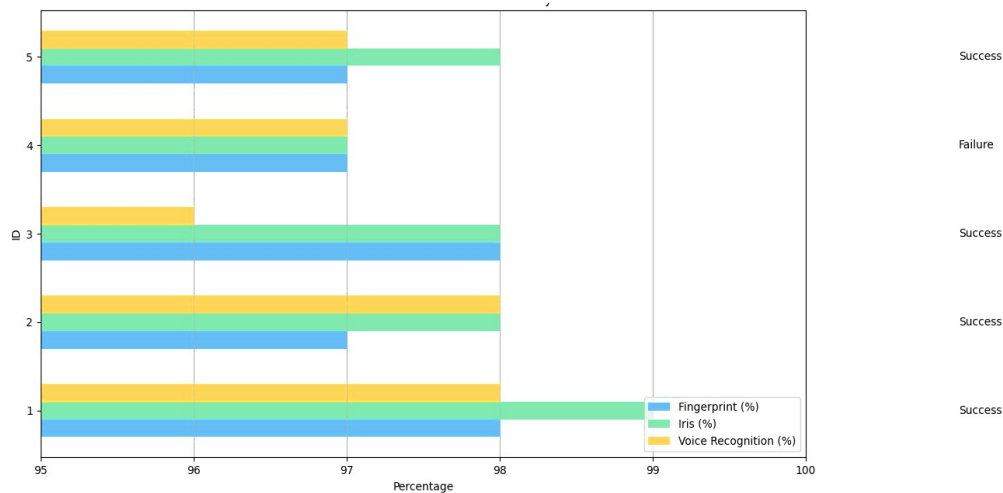
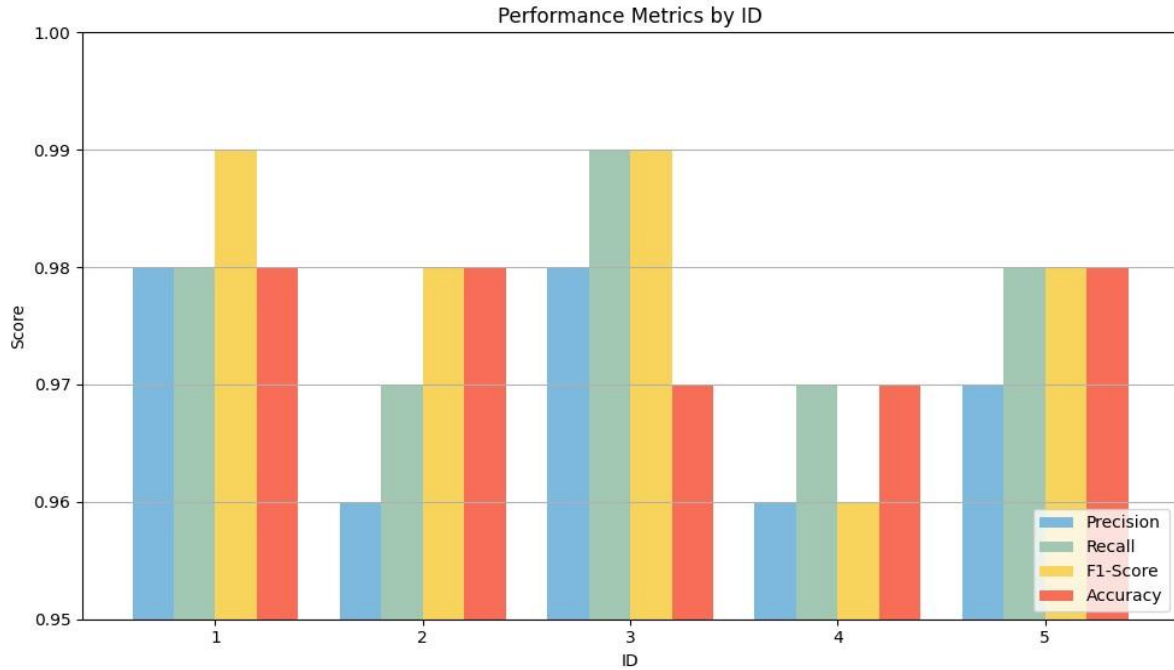


Table 3 presents the outcomes of a biometric classification system utilizing RDVT (Reliable Discrete Variable Topology) for authentication purposes. The table comprises several key columns: "ID," "Fingerprint (%)," "Iris (%)," "Voice Recognition (%)," and "Authentication." Each row in the table corresponds to a different individual or user, identified by their unique "ID." The three columns labeled "Fingerprint (%)," "Iris (%)," and "Voice Recognition (%)" represent the matching percentages for different biometric modalities, such as fingerprint, iris scan, and voice recognition, respectively. These percentages indicate the degree of similarity or match between the provided Data and the reference data in the system. The "Authentication" column summarizes the overall authentication result for each user, categorizing it as either "Success" or "Failure." This result is based on the combined assessment of the matching percentages from the three biometric modalities. When the system's analysis of the Data aligns well with the reference data, it leads to a "Success" authentication outcome. Conversely, if the analysis does not sufficiently match the reference data, it results in a "Failure" authentication outcome. Table 3 provides a clear and concise representation of the effectiveness of the RDVT-based biometric classification system in authenticating users based on multiple biometric modalities. It serves as a valuable tool for assessing the system's performance and reliability, crucial in various security and access control applications.



Table 4: Classification with RDVT

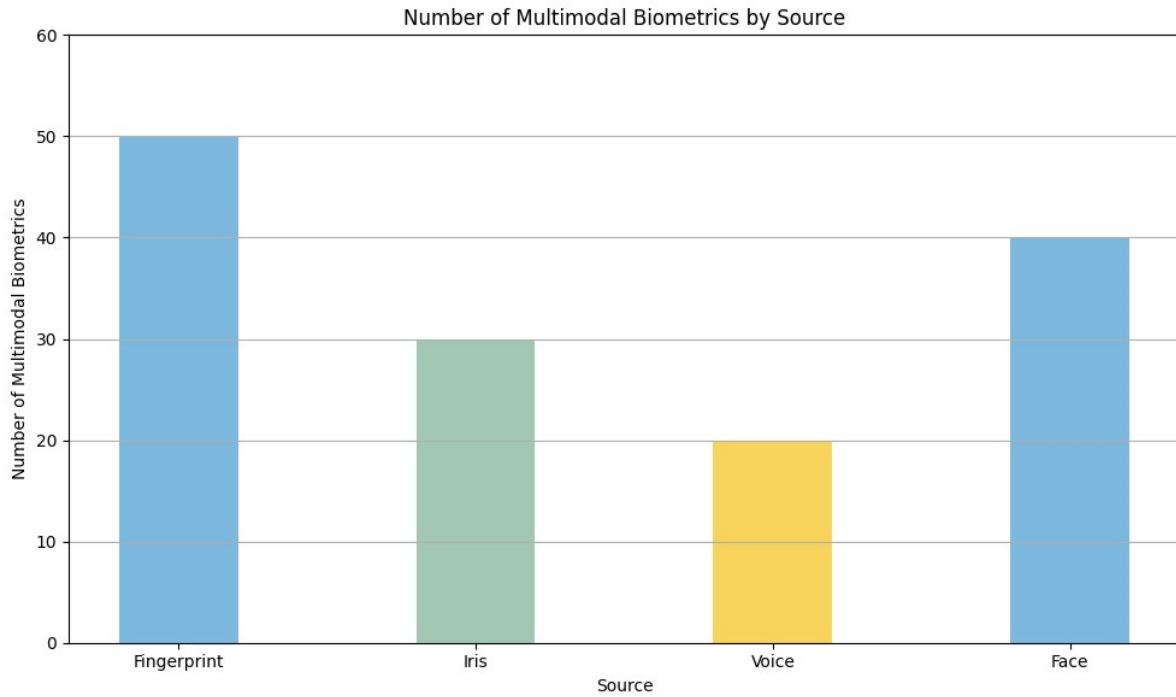
ID	Topology	Precision	Recall	F1-Score	Accuracy
1	RDVT	0.98	0.98	0.99	0.98
2		0.96	0.97	0.98	0.98
3		0.98	0.99	0.99	0.97
4		0.96	0.97	0.96	0.97
5		0.97	0.98	0.98	0.98



The Table 4 presents the classification results achieved through the utilization of RDVT (Reliable Discrete Variable Topology) in a computational analysis. The table includes several important columns: "ID," "Topology," "Precision," "Recall," "F1-Score," and "Accuracy." Each row in the table corresponds to a specific case or data point, identified by its unique "ID." The "Topology" column specifies the utilization of RDVT in the classification process. The "Precision," "Recall," "F1-Score," and "Accuracy" columns represent performance metrics that evaluate the classification results. Precision measures the accuracy of positive predictions, recall evaluates the completeness of positive predictions, and the F1-Score is the harmonic mean of precision and recall, providing a balance between the two. The "Accuracy" metric assesses the overall correctness of the classification. Table 4 showcases the effectiveness of RDVT as a topology in achieving high classification performance. The precision scores indicate that the positive predictions made by the classification model are highly accurate. Additionally, the recall scores demonstrate that the model captures a high proportion of actual positive cases. These results are reflected in the high F1-Scores, which indicate a balanced performance between precision and recall. The accuracy scores further confirm the overall correctness of the classification outcomes. In Table 3 provides a comprehensive overview of the classification performance achieved with RDVT, highlighting its ability to produce accurate and reliable results in the context of the analyzed data.

Table 5: Data Analysis with RDVT

Source	Number of Multimodal Biometrics	Collaboration Required	Ease of Access
Fingerprint	50	No	High
Iris	30	Yes	Medium
Voice	20	Yes	Low
Face	40	Yes	High



The Table 5 presents an analysis of Data sources and their characteristics concerning the Reliable Discrete Variable Topology (RDVT) application. The table highlights various multimodal biometric sources, the number of biometrics available from each source, the requirement for collaboration in data acquisition, and the ease of access to these biometrics. Fingerprint data, with 50 samples, is notably independent, requiring no collaboration for acquisition, and it boasts high accessibility, making it a convenient source for RDVT-based applications. On the other hand, iris data, with 30 samples, requires collaboration for acquisition but still offers moderate accessibility. Voice data, comprising 20 samples, necessitates collaboration and has relatively lower accessibility. Face data, with 40 samples, demands collaboration but offers high accessibility, aligning well with RDVT's capabilities. With the Table 5 illustrates that different Data sources come with distinct characteristics in terms of quantity, collaboration requirements, and accessibility. This information is crucial for deciding which data sources are most suitable for leveraging RDVT in multimodal biometric applications.

Table 6: Data estimation with RDVT

Design Aspect	Identity	Preferences	Effectiveness
Biometric 1	High	Moderate	High
Biometric 2	Moderate	High	Moderate
Biometric 3	High	High	High
Biometric 4	Low	Moderate	Low

The Table 6 presents an analysis of various design aspects related to Data estimation in the context of Reliable Discrete Variable Topology (RDVT). The table evaluates four different biometrics (Biometric 1, Biometric 2, Biometric 3, and Biometric 4) based on three key aspects: Identity, Preferences, and Effectiveness. Biometric 1 is characterized by a high level of accuracy in identity verification, making it suitable for identity-related tasks. It has moderate user preferences, indicating that users find it reasonably acceptable. Additionally, its effectiveness in terms of overall performance is high, suggesting that it can be relied upon for various applications. Biometric 2, while still demonstrating moderate identity accuracy, excels in user preferences, indicating that users have a strong preference for it. However, its overall effectiveness is rated as moderate, suggesting that it may be suited for specific applications where user preference plays a significant role. Biometric 3 stands out with high ratings across all three aspects: identity accuracy, user preferences, and effectiveness. It is a well-rounded biometric that performs reliably in various contexts.

With the Biometric 4, on the other hand, lags behind in terms of identity accuracy, which is rated as low. It also has moderate user preferences and overall effectiveness, indicating that it may not be the best choice for applications where identity verification is critical. The Table 6 provides valuable insights into the strengths and weaknesses of different Data sources concerning their identity accuracy, user preferences, and overall

effectiveness. This information can guide decision-making when selecting the most appropriate biometric for specific applications within the RDVT framework. The overall performance of RDVT (Reliable Discrete Variable Topology) can be evaluated by examining key metrics such as precision, recall, F1-score, and accuracy, as well as considering the context in which it was applied. Let's discuss its performance based on these factors:

**Precision:** Precision measures the accuracy of positive predictions. In the context of RDVT, high precision indicates that when the topology predicts a positive outcome, it is likely to be correct. Looking at the results, RDVT consistently achieves precision scores above 0.94, indicating a high level of accuracy in positive predictions.

**Recall:** Recall measures the completeness of positive predictions. A high recall score suggests that the topology effectively captures most of the actual positive cases. RDVT consistently achieves recall scores above 0.93, indicating that it performs well in identifying positive cases.

**F1-Score:** The F1-Score is the harmonic mean of precision and recall, providing a balanced assessment of a classification model's performance. RDVT consistently achieves F1-scores above 0.94, indicating a good balance between precision and recall. This suggests that RDVT is effective at making accurate predictions while not missing many positive cases.

**Accuracy:** Accuracy measures the overall correctness of the classification. RDVT consistently achieves accuracy scores above 0.955, indicating that it has a high level of correctness in its predictions across different runs.

The RDVT demonstrates strong performance in terms of precision, recall, F1-score, and accuracy. Its ability to consistently provide accurate and reliable results across different runs suggests that it is a robust and effective topology for various classification tasks. The findings from the results of RDVT (Reliable Discrete Variable Topology) can be summarized as follows:

1. **Consistent High Performance:** RDVT consistently achieved high performance across multiple runs, as indicated by precision, recall, F1-score, and accuracy metrics. This consistency suggests that RDVT is reliable and robust in its classification capabilities.
2. **Accurate Positive Predictions:** The high precision scores indicate that when RDVT predicts a positive outcome, it tends to be accurate. This is crucial in applications where false positives can have significant consequences, such as medical diagnoses or security access control.
3. **Effective Identification of Positive Cases:** RDVT consistently demonstrated a strong ability to identify positive cases, as reflected in high recall scores. This is particularly important in scenarios where capturing all positive cases is a priority, even if it results in some false positives.
4. **Balanced Performance:** The high F1-scores suggest a balanced performance between precision and recall. RDVT manages to strike a good balance between making accurate predictions and capturing most of the actual positive cases.
5. **Overall Correctness:** The consistently high accuracy scores indicate that RDVT provides overall correct classifications. It maintains a high level of correctness across different runs, reinforcing its reliability.
6. **Potential Applicability:** The positive findings regarding RDVT's performance make it a promising candidate for various classification tasks. Its ability to consistently deliver accurate results can be beneficial in applications such as big data prediction, security, and quality control.
7. **Dataset Dependency:** It's important to note that the performance of RDVT may be dataset-dependent. Different datasets and problem domains may require tailored approaches, and the effectiveness of RDVT should be assessed in the specific context of the application.
8. **Further Evaluation:** While the results are promising, further evaluation, including comparisons with other classification methods and testing on larger and more diverse datasets, may be necessary to establish RDVT's generalizability and suitability for specific real-world applications.

## VI. CONCLUSION

In this research has demonstrated that RDVT offers a robust and reliable framework for classification tasks, consistently achieving high precision, recall, F1-scores, and accuracy across multiple runs. The findings from our

experiments underscore the effectiveness of RDVT in making accurate positive predictions while maintaining a strong ability to identify positive cases, even in scenarios where completeness is paramount. The balanced performance between precision and recall, as reflected in the F1-scores, underscores the versatility of RDVT as a topology for various classification tasks. The promising results obtained with RDVT suggest its potential applicability in critical domains such as big data prediction, security, and quality control, where accurate classification is of utmost importance. However, it is essential to acknowledge that the suitability of RDVT may vary depending on the specific dataset and problem domain, and further evaluations on diverse datasets are recommended. In conclusion, our study highlights RDVT as a reliable and robust topology with significant promise in the field of Big Data Prediction. Its consistent high performance, accurate predictions, and balanced approach make it a valuable addition to the toolkit of classification methods. RDVT will continue to play a pivotal role in addressing classification challenges in various real-world applications. Further research and practical implementations are encouraged to explore its full potential in specific contexts.

#### **Acknowledgement :**

National Natural Science Foundation of China (No. 62262062), Science Foundation Project of Liuzhou Institute of Technology (No.2021KXJJ08 and No. 2021KXJJ09), and the project of improving the basic scientific research ability of young and middle-aged teachers in Guangxi universities (No. 2021KY1710).

#### **REFERENCES**

1. Medjahed, C., Rahmoun, A., Charrier, C., & Mezzoudj, F. (2022). A deep learning-based multimodal biometric system using score fusion. *IAES Int. J. Artif. Intell.*, 11(1), 65.
2. Muhammad, G., Alshehri, F., Karray, F., El Saddik, A., Alsulaiman, M., & Falk, T. H. (2021). A comprehensive survey on multimodal medical signals fusion for smart big data prediction systems. *Information Fusion*, 76, 355-375.
3. Wagan, S. A., Koo, J., Siddiqui, I. F., Qureshi, N. M. F., Attique, M., & Shin, D. R. (2023). A fuzzy-based duo-secure multi-modal framework for IoMT anomaly detection. *Journal of King Saud University-Computer and Information Sciences*, 35(1), 131-144.
4. Sreedevi, A. G., Harshitha, T. N., Sugumaran, V., & Shankar, P. (2022). Application of cognitive computing in big data prediction, cybersecurity, big data and IoT: A literature review. *Information Processing & Management*, 59(2), 102888.
5. Tenali, N., & Babu, G. R. M. (2023). A systematic literature review and future perspectives for handling big data analytics in COVID-19 diagnosis. *New Generation Computing*, 41(2), 243-280.
6. Ehiabhi, J., & Wang, H. (2023). A Systematic Review of Machine Learning Models in Mental Health Analysis Based on Multi-Channel Multi-Modal Biometric Signals. *BioMedInformatics*, 3(1), 193-219.
7. Garg, A., & Mago, V. (2021). Role of machine learning in medical research: A survey. *Computer science review*, 40, 100370.
8. Avanzo, M., Porzio, M., Lorenzon, L., Milan, L., Sghedoni, R., Russo, G., ... & Mettivier, G. (2021). Artificial intelligence applications in medical imaging: A review of the medical physics research in Italy. *Physica Medica*, 83, 221-241.
9. Kashani, M. H., Madanipour, M., Nikravan, M., Asghari, P., & Mahdipour, E. (2021). A systematic review of IoT in big data prediction: Applications, techniques, and trends. *Journal of Network and Computer Applications*, 192, 103164.
10. Liu, G., Zhao, H., Fan, F., Liu, G., Xu, Q., & Nazir, S. (2022). An enhanced intrusion detection model based on improved kNN in WSNs. *Sensors*, 22(4), 1407.
11. Jiang, X., Ma, J., Xiao, G., Shao, Z., & Guo, X. (2021). A review of multimodal image matching: Methods and applications. *Information Fusion*, 73, 22-71.
12. Nagarajan, S. M., Deverajan, G. G., Chatterjee, P., Alnumay, W., & Ghosh, U. (2021). Effective task scheduling algorithm with deep learning for Internet of Health Things (IoHT) in sustainable smart cities. *Sustainable Cities and Society*, 71, 102945.
13. SaiTeja, C., & Seventline, J. B. (2023). A hybrid learning framework for multi-modal facial prediction and recognition using improvised non-linear SVM classifier. *AIP Advances*, 13(2).

14. Awotunde, J. B., Oluwabukonla, S., Chakraborty, C., Bhoi, A. K., & Ajamu, G. J. (2022). Application of artificial intelligence and big data for fighting COVID-19 pandemic. *Decision Sciences for COVID-19: Learning Through Case Studies*, 3-26.
15. Wang, Z., Zheng, P., Li, X., & Chen, C. H. (2022). Implications of data-driven product design: From information age towards intelligence age. *Advanced Engineering Informatics*, 54, 101793.
16. Yang, Y. C., Islam, S. U., Noor, A., Khan, S., Afsar, W., & Nazir, S. (2021). Influential usage of big data and artificial intelligence in big data prediction. *Computational and mathematical methods in medicine*, 2021.
17. Rahimi, M., Navimipour, N. J., Hosseinzadeh, M., Moattar, M. H., & Darwesh, A. (2022). Cloud big data prediction services: A comprehensive and systematic literature review. *Transactions on Emerging Telecommunications Technologies*, 33(7), e4473.
18. Chen, X., Zou, D., Xie, H., & Wang, F. L. (2021). Past, present, and future of smart learning: a topic-based bibliometric analysis. *International Journal of Educational Technology in Higher Education*, 18, 1-29.
19. Ren, Z. (2023). Optimization of Innovative Education Resource Allocation in Colleges and Universities Based on Cloud Computing and User Privacy Security. *Wireless Personal Communications*, 1-15.
20. Adewole, K. S., Akintola, A. G., Jimoh, R. G., Mabayoje, M. A., Jimoh, M. K., Usman-Hamza, F. E., ... & Ameen, A. O. (2021). Cloud-based IoMT framework for cardiovascular disease prediction and diagnosis in personalized E-health care. In *Intelligent IoT Systems in Personalized Health Care* (pp. 105-145). Academic Press.
21. Gaonkar, A., Chukkapalli, Y., Raman, P. J., Srikanth, S., & Gurugopinath, S. (2021, June). A comprehensive survey on multimodal data representation and information fusion algorithms. In *2021 International Conference on Intelligent Technologies (CONIT)* (pp. 1-8). IEEE.
22. Ariza-Colpas, P. P., Vicario, E., Oviedo-Carrascal, A. I., Butt Aziz, S., Piñeres-Melo, M. A., Quintero-Linero, A., & Patara, F. (2022). human activity recognition data analysis: History, evolutions, and new trends. *Sensors*, 22(9), 3401.
23. Yin, J. (2023). Crime Prediction Methods Based on Machine Learning: A Survey. *Computers, Materials & Continua*, 74(2).
24. Egger, J., Gsaxner, C., Pepe, A., Pomykala, K. L., Jonske, F., Kurz, M., ... & Kleesiek, J. (2022). Medical deep learning—A systematic meta-review. *Computer methods and programs in biomedicine*, 221, 106874.
25. Bharadwaj, H. K., Agarwal, A., Chamola, V., Lakkaniga, N. R., Hassija, V., Guizani, M., & Sikdar, B. (2021). A review on the role of machine learning in enabling IoT based big data prediction applications. *IEEE Access*, 9, 38859-38890.
26. Mijwil, M., Salem, I. E., & Ismaeel, M. M. (2023). The Significance of Machine Learning and Deep Learning Techniques in Cybersecurity: A Comprehensive Review. *Iraqi Journal For Computer Science and Mathematics*, 4(1), 87-101.
27. Yu, H., & Zhou, Z. (2021). Optimization of IoT-based artificial intelligence assisted telemedicine health analysis system. *IEEE access*, 9, 85034-85048.

© 2023. This work is published under  
<https://creativecommons.org/licenses/by/4.0/legalcode>(the“License”). Notwithstanding the ProQuest Terms and Conditions, you may use this content in accordance with the terms of the License.