BIG DATA-MODEL INTEGRATION AND ARTIFICIAL INTELLIGENCE-BASED VECTOR-BORNE DISEASE PREDICTION

Muhammad Hamza Imtiaz and Dr. Fadia Shah

ABSTRACT

An intelligent system for diagnosing vector-borne diseases using machine learning techniques is proposed in this thesis. The system aims to address the time-consuming, costly, and limited traditional mosquito surveys. Mosquito monitoring data is crucial in identifying increasing pesticide resistance, supporting effective insecticide intervention programs, and modeling vector-borne disease transmission. The proposed system uses classification algorithms such as Logistic Regression, Neural Network, K-Nearest Neighbor, Naïve Bayes, XG Boost, and Decision Tree, and feature selection algorithms such as Relief, Minimal Redundancy Maximal Relevance, Least Absolute Shrinkage Selection Operator, and Local Learning. A novel fast conditional mutual information feature selection algorithm (FCMIM) is also proposed to address feature selection problems. The system's performance is evaluated using the leave-one-subject-out cross-validation method and performance metrics. The experimental results show that the proposed system achieves good accuracy compared to previously proposed methods and can be easily implemented in healthcare for the identification of vector-borne diseases. The system's feasibility with Support Vector Machine classifiers and the proposed FCMIM feature selection algorithm demonstrates the potential for developing a highly intelligent system for diagnosing vector-borne disease. This thesis highlights the importance of machine learning techniques in the identification and diagnosis of vector-borne diseases, leading to better patient outcomes and improved healthcare delivery. Further research and development of similar machine learning-based systems could have significant implications for the field of Lyme disease.

Keywords: Model based testing, Artificial Intelligence, Data Analytics, Prediction, Classification Techniques, Big Data.

BIG DATA-MODEL INTEGRATION AND ARTIFICIAL INTELLIGENCE-BASED VECTOR-BORNE DISEASE PREDICTION

Huria Ali and Dr. Muhammad Imran

ABSTRACT

Plant disease detection and classification is an area of research that plays a significant role in the prosperity of the agricultural sector. Accurately identifying plant diseases is crucial as it enables the implementation of appropriate precautions and measures to safeguard the affected plant. Incorrect identification, on the other hand, can result in the application of ineffective measures, leading to significant wastage of time and money. In this particular study, the YOLO-V5 object detection network is proposed for the detection and classification of various plants' (fruits & vegetables) diseases. In the initial stage, the dataset images go through a pre-processing phase. This phase involves class balancing & data augmentation and data annotation. In the next stage, the pre-processed images are fed into five YOLO-V5 network variants (n, s, m, l & x) for class predictions. The dataset utilized in this study is the PlantVillage dataset. It contains plant images focused on fruits and vegetables. The five YOLO-V5 variants are compared and analyzed. According to the results, YOLO-V5x variant outperforms all other variants with a precision of 98.5%, recall of 96.4%, F1-score of 97.6% and mAP (0.5 - 0.95) of 88.1%.

Keywords: YOLO-V5, Deep Learning (DL), Pre-processing, Class balancing, Data Augmentation, Data Annotation.

A GAME-BASED IMAGE-CAPTCHA GENERATION AND SECURITY ANALYSIS

Hafiz Umar Farooq and Dr. Danish Mehmood

ABSTRACT

The Internet is a vital source of interaction for everyone today. It makes no difference what your profession, status, gender, or age is. With the abundance of a diverse range of smart devices used globally, the threat of bot attacks has increased. Users access the Internet for various needs, such as entertainment, education, commerce, and data exchange. It has tremendously increased the number of users and connected devices. The completely automated public Turing test to tell the computer and human apart (CAPTCHA) is a fully automated Turing test to differentiate a machine from a person and safeguard internet services from spam and bot assaults. With evolving technology and time, multiple CAPTCHAs are devised; however, most of these can easily be broken. For this reason, noises may be added systematically to an existing CAPTCHA to increase its security. However, the distortion levels make human identification unpleasant and impossible. On the other hand, traditional CAPTCHAs sometimes need the use of keyboards and mouse, restricting current smart hand-held devices like smartphones. However, some of the CAPTCHAs are weaker, allowing hackers to intrude the user accounts. Only a few bot programs have got the ability to solve CAPTCHAs. They automatically complete the registration process and cause traffic congestion, minimizing the system's performance. This thesis describes the image CAPTCHA that compares the strengths and weaknesses of previously surveyed papers rather than relying on complicated alphabet inputs. We provide a revolutionary Game-based CAPTCHA scheme, "GAPTCHA," based on a famous game called "Ludo snake and ladder game." It uses a simple keyboard, mouse, and finger inputs, is language agnostic, increases security without bothering users, and makes CAPTCHAs more engaging and pleasant yet strong against bot attacks. In the results section, 97% of users had significant time to solve the proposed GAPTCHA correctly. However, only 3% of the people got it wrong and said it is not entertaining most of the time. We also asked the users about using the proposed GAPTCHA again. 100% of the users said they would like to reuse this proposed GAPTCHA, which is a positive result for its successful user experience.

Keywords: CAPTCHA, Game-Based CAPTCHA.

IMAGE DEHAZING AND VISIBILITY ENHANCEMENT USING DEEP LEARNING

Muhammad Baqar Kazmi and Dr.Muhammad Imran

ABSTRACT

Single image dehazing has always been a challenging problem in the field of computer vision. Traditional image dehazing methods use manual features. With the development of artificial intelligence, the dehazing method based on deep learning has developed rapidly. In this paper, we propose a novel image dehazing approach called NIN-DehazeNet for single image. This method estimates the transmission map by NIN-DehazeNet combining Network-in-Network with MSCNN(Single Image Dehazing via Multi-Scale Convolutional Neural Networks). In the test stage, we estimate the transmission map of the input hazy image based on the trained model, and then generate the dehazed image using the estimated atmospheric light and computed transmission map. Extensive experiments have shown that the proposed algorithm over performance traditional methods.

Keywords: Dark Channel Prior, Structural Similarity (SSIM), Peak Signal-to-Noise Ratio (PSNR), DeHazing, Feature Extraction, Deep Learning (DL).

AN ADAPTIVE NETWORK BASED ON PCA AND SUPPORT VECTOR MACHINES BY USING CNN ON FACIAL EMOTION RECOGNITION

Humayoon Siddique and Dr.Fadia Shah

ABSTRACT

FER is the process of detecting and interpreting human emotions (HM) from facial cues. It is a rapidly growing field of research with a wide range of applications, including psychology, marketing, education, and human-computer interaction. The diversity of facial expressions among people and cultures is one of the biggest obstacles to facial expression identification. People can express the same feeling in a variety of ways, and cultural standards on emotional expression may vary. Another issue is the dearth of high-quality annotated data sets, which are necessary for developing and testing machine learning face expression detection algorithms. There are several different approaches to Facial Expression Recognition (FER), including rule-based methods, which rely on hand-crafted rules to identify specific facial features, and machine learning-based methods, which learn to recognize facial expressions from annotated data sets. Deep learning techniques, in particular, have achieved state-of-the-art performance in a number of face emotion identification tests. In this study, using three datasets (Cohn-Kanade (CK+), Facial Expression Recognition-Fer-2013, and NCAER-S) and employing specific pre-processing techniques to produce photographs in fine form, we then employ the Deep Learning (DL) model named ResNet-50 to achieve high performance and accuracy.

Keywords: Facial Expression Recognition (FER), Emotion Recognition (ER), Computer Vision (CV), Deep Learning (DL), National Council of Applied Economic Research (NCAER).

AUTOMATIC OUTDOOR BULLET IMPACT DETECTION & GROUP CALCULATION SYSTEM

Naveed Ahmad and Dr. Shahzad Lattif

ABSTRACT

The Automatic Outdoor Bullet Impact Detection & Group Calculation System is designed for use in military shooting exercises. Its purpose is to detect when bullets have been fired and to accurately calculate the location of the bullet impacts. This information can then be used to assess the accuracy of the shooters and to provide feedback on their performance. The system is intended to be used outdoors, and it is designed to be automatic, requiring minimal input from the user. The system is designed to automatically detect and identify bullet impacts on a target, and to calculate the group of firers responsible for each impact. The system is tested using a 50 by 50cm sheet as the study target, and evaluated in both laboratory and military shooting range environments. In order to develop this system, reviewed the existing methodologies, and identified the limitations of existing approaches. In order to overcome these limitations, the research propose and evaluate new methods for target image preprocessing, bullet hole identification, and firer group calculation. The research incorporated a range of image processing techniques to overcome the limitations of the current approaches. These methods encompass dilatation, target ring recognition, perspective translation, and morphological image processing. To accurately identify the bullet hole locations and determine the corresponding firer group for each impact, the research adopted the Mask R-CNN method. The methodology involves preprocessing, impact detection, edge detection, and result evaluation. Through experimentation with 416 shooting sessions and 2759 bullet impacts, the study demonstrates the effectiveness of the proposed method. Notably, the ResNet-50 model surpasses the ResNet-101 model, achieving impressive accuracy, precision, and recall scores of 98.8%, 99.5%, and 99.3%, respectively. Edge detection and bullet impact count exhibit 100% effectiveness. The proposed system's performance was thoroughly assessed using authentic data obtained from military shooting exercises. A comparative analysis conducted between the results of the proposed approach and existing methods to establish its effectiveness.

Keywords: Contour Detection, Image Processing, , Group Calculation, Open Environment Bullet impact detection, Target Score Calculation, Bullet Impact Location detection.

ARTIFICIAL INTELLIGENCE-BASED FRAUD DETECTION TECHNIQUE USING UNSUPERVISED MACHINE LEARNING FOR REAL ESTATE PROPERTY INSURANCE

Saman Ikram Abbasi and Dr.Fadia Shah

ABSTRACT

Insurance fraud is a pressing issue, and traditional prevention methods have proven ineffective. This study explores how Artificial Intelligence (AI), specifically unsupervised machine learning, can address property fraud issues in the real estate property insurance industry. Utilizing a dataset of 1,070,994 property records with 32 structured data fields, including property location, value, and relevant features, the analysis employs unsupervised learning techniques such as clustering algorithms. Among these, Principal Component Analysis (PCA) and autoencoder simplify the dataset while retaining essential information, allowing for efficient identification of potential fraud cases. The study goes beyond mere identification, incorporating the generation of fraud scores. Notably, variables like FULLVALL, AVLAND, AVTOT, and their ratios exhibit variations. For instance, FULLVALL/BLDAREA and AVLAND/BLDVOL in specific groups show higher z-scores. In the results analysis, an in-depth exploration of the top ten properties exhibiting the highest fraud scores was conducted using unsupervised machine learning. This investigation aimed to illuminate factors contributing to the anomalous scores, with specific emphasis on variables such as LTFRONT and LTDEPTH, which demonstrated unusual patterns characterized by either extremely low or high values. The identified patterns strongly suggested the presence of potential fraudulent activities within the dataset. The examination of notable irregularities, particularly within variables like LTFRONT and LTDEPTH, served as a crucial indicator of potential fraudulent records or clerical errors. To further investigate and enhance our understanding, a subsequent analysis was conducted on z-scaled variables associated with these top records. These findings contribute valuable information for refining and strengthening the effectiveness of fraud detection mechanisms in the dataset.

Keywords: Insurance Fraud Detection, Real Estate Fraud, Artificial Intelligence, Fraud Detection, Machine Learning, Unsupervised Learning.

SOFTWARE RELIABILITY ANALYSIS USING PREDICTION MODEL

Ali Imran Haidry and Dr. Tahir Jameel

ABSTRACT

Software Reliability analysis plays a vital role in ensuring the quality and reliability of software systems. To achieve the desired reliability of a software product, early reliability prediction in the software development life cycle is beneficial for effective resource allocation, risk assessment, and decision-making. These early decisions have a twofold benefit: they directly enhance software reliability and reduce rework costs associated with bug fixes at later stages. In this research, different approaches to software reliability prediction are explored, and Neufelder's shortcut model is applied to real projects to evaluate its effectiveness and usefulness. The prediction models are based on historical failure data, capturing complex dependencies and relationships among various factors that directly influence software reliability. These models have an advantage over reliability growth estimation models, which are typically used after the testing phase, as they can predict reliability early in the project life cycle, enabling corrective decisions proactively to achieve the desired reliability more deliberately. Real software metrics were collected for Neufelder's shortcut model across different projects to predict defect density and software reliability, and their effectiveness was evaluated using designer assessments and growth estimations based on the failure data of the software products. The results demonstrate that software reliability prediction models can be valuable for developers, QA engineers, reliability engineers, and customers for the pre-assessment of the desired product's reliability. Overall, this research contributes to advancements in software reliability analysis, empowering organizations to build more reliable and robust software systems.

Keywords: Shortcut Model, Defect density, Reliability prediction.

LEARNING BASED FRAMEWORK FOR GPS SPOOFING ATTACK DETECTION IN UNMANNED AERIAL VEHICLE

Aziz Ur Rehman Badar and Dr. Danish Mehmood

ABSTRACT

Unmanned Aerial Vehicles (UAVs) are used commonly these days in different infrastructures like transportation, delivery, monitoring, and aviation. One of the major security threats is the GPS Spoofing attack in which the attacker manages to fool the GPS receiver with fake signals and take over the control of the UAV which can lead to the capture of the UAV or destroying it in some cases. There are many solutions proposed over the years to detect spoofed GPS signals. While GPS spoofing attack detection using machine learning has seen significant advancements in recent years, there are still some challenges and limitations associated with the state-of-the-art work in this field like imbalanced datasets, lack of optimal feature selection, and accuracy of attack detection in resource-constrained environments for deployment. In this thesis, we proposed a framework that uses a combination of different methods of feature selection and Neural Networks (NN) for GPS spoofing attack detection. This study focuses on the difficulties of identifying GPS spoofing attacks in Unmanned Aerial Vehicles (UAVs) by presenting a thorough solution that integrates feature selection techniques and sophisticated neural network structures. The work concentrates on addressing the problem of unbalanced datasets by employing rigorous oversampling strategies. The characteristics that have been chosen during this procedure are subsequently sent to the next layers of a hybrid neural network, which combines Convolutional Neural Network (CNN) and Bidirectional Long Short-Term Memory (BiLSTM) components. The hybrid model proposed for detecting GPS spoofing attacks demonstrates considerable gains in terms of prediction accuracy, true positive and false positive rates, as well as F1-Score and Recall values. The ANOVA + CNNBiLSTM hybrid model proved to be the most successful design, achieving remarkable results with a precision of 98.84%, accuracy of 99.25%, F1 Score of 99.26%, and recall of 99.69%.

Keywords: GPS Spoofing, UAV Attacks, Neural Networks, correlated features, Bi-LSTM, deep learning models, sensors data

DEEPFAKE IDENTIFICATION OF SOCIAL MEDIA IMAGES AND VIDEOS

Laraib Razzaq and Dr. Muhammad Imran

ABSTRACT

In the age of social media, deepfake technology has grown to be a serious worry since it makes it possible to manipulate media to produce incredibly realistic images. Researchers have used Generative Adversarial Networks (GANs) for deepfake identification to overcome this problem. The Celeb DF v2 dataset, which includes a varied collection of actual and deepfake celebrity photos and videos, is used in the GAN-based approach to train a discriminator network. By recognizing minute visual clues and artefacts specific to deepfakes, the discriminator network learns to differentiate between genuine and altered data. The generator network seeks to build deepfakes that are indistinguishable from authentic content through an adversarial training procedure, testing the discriminator's ability to recognize fake information. The objective of this deepfake identification study on the Celeb DF v2 dataset is to enhance the model's generalization to previously unknown deepfake variations and detection of more advanced deepfake approaches. The GAN architecture incorporates methods like deep neural networks, convolutional neural networks, and recurrent neural networks to improve model robustness and detection accuracy. Additionally, attempts are being undertaken to improve the GAN-based deepfake identification models' interpretability and explain ability so that users may comprehend the justification for the models' judgments. The GAN Accuracy Cycle, which consisted of consecutive repetitions with increasing accuracy values, peaked at 96.59 after ten iterations. Experimental results demonstrated that the proposed framework performed well than State-of-art techniques.

Keywords: GAN, Celeb DF V2, Deepfake Identification, Discriminator, Generalization.

RESOURCE OPTIMIZATION IN COGNITIVE RADIO NETWORKS (CRNS) USING EVOLUTIONARY TECHINQUES

Ahmad Bilal and Dr. Shahzad Latif

ABSTRACT

The spectrum demand is increasing exponentially due to the inventions of new technologies. The enormous spectrum demand creates spectrum scarcity problem. Cognitive radio (CR) technology is advance gadget to deal with unavailability of spectrum. The primary users referred as Mobile users (MU) and secondary users as Device users (DU) which utilize available resources but only MU have permission for using resources at any time they want, and DU can use resources only in the idle time when MU are not using the spectrum. MU can take the resources from DU at any time they want. Due to which interference is occurred which becomes a problem for sharing of resources in CR networks. The main objective in CR communication is to optimize available resources (throughput) so that enough resources could be available for DU with having minimum interference factor. D2D communication system for DUs in a cognitive (MU) network is used for reusing the channels for unlicensed band. D2D communication system, which is controlled by the mobile network shares the alike spectrum, as mobile communication is increasing the spectrum's utility. Mobile network is not harmed because of D2D communication system. IMT-Advanced systems, like 5G and WiMAX, permit D2D communication system as dedicated resource or shared resource, may be used for D2D communication system in mobile network to enhance spectral ability. There are many techniques to optimize the spectrum resources. Evolutionary computing is good to deal with optimization problems. The techniques proposed for extracting solution of the problem are known as Modified Non-Dominated Sorting Genetic Algorithm (MNSGA) and Modified Whale Optimization Algorithm (MWOA). The validity of proposed algorithms is verified through simulations and comparisons with other literature techniques. The simulation results showed that the proposed algorithms outperformed in comparison to results of the existing techniques.

Keywords: Resource Optimization, Cognitive Radio Networks, Evolutionary Techniques.

VISIBILITY ENHANCEMENT OF SCENE IMAGE DEGRADED BY FOGGY WEATHER CONDITION

Gulfam Zahra and Dr. Muhammad Imran

ABSTRACT

In recent years, machine vision, computer vision, and image processing have played a significant role in our daily lives. Images captured in adverse weather condition have reduced visibility due to the presence of fog, haze, snow, and mist. Various outdoor applications faced challenges while dealing with images captured during foggy and hazy weather, such as surveillance, object tracking, transportation, and object identification. Therefore, image defogging and dehazing is an important field for computer vision applications. To improve visibility and adjust contrast, a number of visibility enhancement techniques have been proposed in the past. However, these techniques suffer from limitations that place strong obstacles to real-world outdoor computer vision applications. Hence, there is a dire need for a reliable technique to enhance the visibility of foggy and hazy images. This research work proposes a new framework to defog and dehaze the image based on Conditional Generative Adversarial Network (CGAN). The goal of this research is to effectively remove fog, enhance visibility, adjust contrast, and restore true color details of bad weather degraded foggy and hazy images. The proposed framework consists of two networks: the generator network generates fog-free images, and the discriminator network distinguishes between the restored image and the original fog-free image. This framework directly estimates the transmission map, atmospheric light, and scene radiance. Experiments were conducted on FRIDA and Haze datasets, and five performance metrics (PSNR, SSIM, e, r, and 6) were used to evaluate the proposed method's performance. The experimental results demonstrated that the proposed framework can directly remove fog and enhance the visibility of foggy and hazy images, outperforming state-of-the-art techniques.

Keywords: Visibility, Bad Weather, Degraded Images, Generative Adversarial Network, Transmission Map, Artifacts, Image Restoration.

AN RPL-ENABLED SECURITY FRAMEWORK FOR MITIGATING INTERNAL ATTACKS FOR LOW POWER AND LOSSY NETWORKS-BASED IOT

Muhammad Hassan and Dr. Noshina Tariq

ABSTRACT

Internet of Things (IoT) encouraged another age of development by interconnecting advanced devices, which makes the existence of human life helpful and more attractive. These advanced devices are generally termed as Low power and Lossy Networks (LLNs) for offering assistance in different areas of application. The LLNs are a type of networks in which the inter-connected devices are severely resource constrained in terms of processing, memory and power. They are characterized by high failure rate, low information rate, and instability in correspondence links. This thesis is working on resource-constrained devices. The Engineering Task Force (IETF) has planned a lightweight IPv6 network routing convention called Routing Protocol for LLNs (RPL), which works with the correspondence between LLN-based network devices. Nonetheless, the RPL routing protocol is highly at risk against an enormous variety of both outer and insider attacks to cause wrecking and cataclysmic impacts. RPL offers resistance against outer attacks but it very vulnerable to many internal attacks. The extent of this thesis rotates around inside attacks, where nodes are a genuine part of the network. In this way, trust-based security helps build up a safe and trustworthy environment against these types of attacks. However, existing trust based security mechanisms have high energy consumption issue, which causes reduction in the performance of LLNs because of complex calculation at the node level. Therefore, this thesis presents a novel trust-based mechanism "FTrust-RPL" which not only secures the LLN devices from insider attacks but also utilizes less energy comparatively. FTrust-RPL mechanism provides a secure system especially against Sybil attack by checking the trust of nodes based on their forwarding or sending data behavior. This thesis shifts the complex trust related computations to the upper layer known as Fog layer to save the computational and energy resource at the node level. The proposed system is evaluated using cooja network simulator. Different parameters are used for evaluation such as detection rate, energy consumption, end to end delay and number of messages overhead. It can be observed from results that proposed model performs better than the existing techniques. Our mechanism demonstrates a superior performance over others in detecting and isolating sybil attacks. The results depict that FTrust-RPL detected and isolated more malicious nodes than other techniques in the same time lapse. The sybil attack detection rate increases by 11.12%, energy consumption reduced by 21%, isolation latency reduces by 39% with respect to time while end to end delay is reduced by 4.18% and no of control messages increases by 25%.
Keywords: LLNS, Internet of Things, Nodes, trust, C-trust, IPV6, RP.

AGILE SOFTWARE DEVELOPMENT EFFORT ESTIMATION USING DEEP LEARNING TECHNIQUES

Danish Nawaz and Dr. Shahzad Latif

ABSTRACT

In Software engineering the role of estimation is vital for agile software development. Its cost effect and the required effort consequently influence the overall software development success. An expert-based system, algorithmic base methods, analogy-based approach; the error margin is quite significant, which express the software project in the running over budget and danger of delays. This study analyses the agile software development effort estimation and its effectiveness with its associated problems. For better estimates, this study is an alternative method with data mining on the historical data. Investigation in this research suggests that performing the technique of machine learning like SVM (support vector machine) and CNN (convolutional neural network) can enhance the effort estimation. The generated model is then tested by using the five folds of cross-validation. The SVM modelling framework consists of two parts: training and data set training, with the goal of producing a model that can predict target value with a given data attribute. A data flow model is presented below, along with a sample of data for error estimates. In the end, it was evaluated by using precision, classification accuracy, AUC (accuracy) and Recall. The results of the estimation are compared with the assessment. All applied techniques are successful in getting better results. To meet future requirements, agile software development it proved to be the better option. According to the thorough SLR, agile software engineering is having difficulty scaling to large-scale projects. It is also having difficulty estimating effort and expense. GSD (global software development) is used to solve the vast majority of agile difficulties. SCRUM, scrum of scrum, and lean are now the most popular agile approaches on the market. If machine learning techniques can be applied to agile domains, they can play a critical role in agile project work estimates. So far, CNN, KNN, LSTM, KSVM, and SVM have been the most widely utilized data science methods in software engineering areas.

Keywords: Agile, SLR, Effort Estimation, Statistical Analysis, CNN, SVM, KNN, LSTM, Accuracy, Precision, Agile Software Development, Estimation.

PATIENT PAIN DETECTION FROM FACIAL EXPRESSIONS USING DEEP LEARNING

Mehboob Ali and Dr. Muhammad Imran

ABSTRACT

With the advancement in deep learning techniques and lower cost of computation resources, there have been plenty of approaches towards FER. The more complex architectures based on deep networks enable to operate on large datasets and in variety of applications. The deep learning approach has improved the accuracy and scalability of the models as compared to digital image processing techniques and computer vision algorithms. But still there is room for improvements in accuracy and performance especially for small devices with less memory and computational power. If we took a basic deep learning approach, still the computation it requires made it impractical for low processing applications. There are plenty of models proposed to overcome the above challenges. Some researchers have tried to combine hand craft features with CNN. Some have proposed hybrid computer vision and deep learning approaches. Some have proposed Generative Adversarial Model in order to train the models well for better accuracy. Some have tried to generalize the model for scalability. The contribution of this study is to comprehensive analyze different deep learning approaches & identify areas where further improvement is required to balance the accuracy and performance of current algorithms. In this study we also relate different variations with each other to choose the best algorithm and propose a model to overcome the drawback of existing models.

Keywords: Convolutional Neural Network, Deep Learning, Pain detection, Expression Recognition, Classification.

FRAMEWORK FOR PATIENT DATA PRIVACY AND SECURITY IN M-HEALTH WITH COMPLIANCE TO HIPPA

Almas Tabsaam and Dr. Isma Masood

ABSTRACT

Mobile Health (mHealth) applications play an important role in the field of health care and provide services to the patient. The usage of mHealth apps rapidly increased in the era of mobile phones because it's very easy for the patient to manage their health (vitals, medication, and appointments) with their physician. A telehealth term is also practiced for remote patient monitoring, to contact the physician from a remote place, where the physician is not able to come to visit the patient. However, mHealth apps used the sensitive patient data that carried from mHealth apps to the server-side database and used the transport layer for communication of data between client-server. By using sensitive data and communication channels, mHealth has potential risks of patient data privacy and security. mHealth apps are mostly dependent on a remote server over the internet. Communication through public infrastructure can be observed, modified, and redirected. Numerous mHealth applications do not utilize any encryption procedure between client-server that makes an imminent danger for the patient information security and data penetration and are additionally conceivable. These issues need to be addressed and get the researcher's attention because it's necessary for maintaining the privacy and security of mHealth apps and increase the current ratio of end-users. The aim of this study is to identify the transport layer security issues relevant to client-server of mHealth apps. To test, a subset of health care android and iOS apps collected from the Play Store and App Store, and to evaluate the client security, a Charles Proxy tool selected to test client applications. We proposed a framework to develop the mhealth application and apply the security parameters to reduce the risk of data breach. At the end of the thesis, we presented the principal results of how many apps have vulnerabilities concerned on transport security and compare our results for transport layer of SSL and TLS certificates and chipper suit strength to show performance of our proposed framework.

OPTIMAL FEATURE SELECTION FOR IMPROVED PREDICTION IN CARDIOVASCULAR DISEASE

Farman Ul Haq and Dr. Noshin Tariq

ABSTRACT

According to the World Health Organization (WHO), Cardiovascular Diseases (CVDs) are a leading cause of death worldwide, accounting for a significant number of annual fatalities. Machine Learning (ML) techniques have emerged as valuable tools in the healthcare industry, offering opportunities to develop systems that aid in the diagnosis of heart disease. By accurately predicting CVD, lives can be saved, and the decision-making process of doctors can be improved while reducing time and costs. Numerous studies have utilized ML approaches to predict CVD, with a focus on feature selection to identify the most relevant attributes from the dataset. The objective of this study is to identify an optimal subset of features and ML techniques that enhance the accuracy of CVD classification. Various feature selection techniques, including Chi-square, Info Gain, Forward Selection, and Backward Elimination, were employed to obtain a subset of 8 features. ML models such as Random Forest, XGBoost, Decision Tree, and Logistic Regression were trained on this subset. Among the models, XGBoost yielded the highest results when using the feature subset extracted through Info Gain. It achieved an accuracy, recall, f-measure, and precision of 99%, 100%, 99%, and 99%, respectively. Additionally, Principal Component Analysis (PCA) was applied to further reduce the dimensions of the feature subset from 8 to 6 attributes. Once again, XGBoost outperformed the other models, achieving an accuracy, recall, f-measure, and precision of 98%, 100%, 98%, and 97% on the feature subset obtained through Chi-square and Forward Selection methods. This research highlights the significance of ML and feature selection techniques in improving the accuracy of CVD classification. By identifying the optimal features subset and employing ML models, accurate predictions can be made for early detection and diagnosis of CVD.

Keywords: Machine Learning, Feature Selection, Optimal Features, CVD.

AUTOMATIC IDENTIFICATION OF TWITTER EYEWITNESS MESSAGES DURING DISASTER

Muhammad Mubasher Abbasi and Dr.Muhammad Imran

ABSTRACT

Social media applications such as Facebook, Twitter, and Instagram provide valuable information that can be utilized by rescue and relief agencies. People often share their experiences on these platforms, and this information can help agencies plan their resources effectively during disasters. Accurate information about ongoing disasters is crucial for these agencies to respond appropriately. Eyewitness reports are particularly valuable as they provide authentic and real-time information. This research aims to classify Twitter reports into three categories: direct eyewitness reports, indirect eyewitness reports, and vulnerable eyewitness reports. Additionally, the characteristics of eyewitnesses will be studied. Machine learning techniques, combined with domain expert features, will be employed to classify Twitter tweets accurately and determine the type of eyewitness report they represent. By classifying the tweets, rescue and relief agencies can prioritize their response based on the nature of the information provided. Direct eyewitness reports, for example, may be given higher priority as they provide firsthand information about the disaster. The study will contribute to the development of effective strategies for utilizing social media data in disaster response and management.

Keywords: Social Media, Twitter, Eyewitness Reports, Machine Learning, Classification, Disaster Response.

AN ADVANCED APPROACH TO DETECT SYBIL ATTACKS

Shayan-E-Ali and Dr. Noshina Tariq

ABSTRACT

Sensitive applications, such as healthcare and medical services, require secure and reliable transmission of data. However, these applications are vulnerable to attacks, such as the Sybil attack, where false nodes are created to disrupt the network. In healthcare, where accurate and reliable data is crucial, it is essential to identify and isolate these false nodes to ensure the integrity of the system. This is particularly important in the context of life-threatening diseases like COVID-19, where devices connected to the Internet of Medical Things (IoMT) need to be trustworthy. To address these security concerns, this study proposes a blockchain-based trust management framework for detecting and isolating Sybil nodes in IoMT networks. The framework utilizes blockchain technology to securely record transactions and detect duplicate identities in case of a Sybil attack. It also incorporates a Fuzzy Logic trust model to effectively manage security issues in the IoMT network. The proposed mechanism is implemented and evaluated using the Cooja network simulator, considering quality of service (QoS) parameters. The results show that the proposed framework consumes 25.43% less energy compared to existing state-of-the-art approaches. It detects attacks with a higher accuracy of 12.54% compared to existing work and demonstrates a 37.85% improvement in trust computations. Moreover, the framework achieves a higher packet delivery ratio of 17.40% and better throughput, outperforming existing approaches by 13.04%. The blockchain-based trust management framework offers enhanced security and reliability for IoMT networks, ensuring the integrity and trustworthiness of sensitive healthcare applications.

AI-BASED FRAMEWORK FOR EARLY DETECTION OF COVID-19 USING X-RAY IMAGE

Kainat Khero and Dr. Muhammad Usman

ABSTRACT

This study explores the pivotal role of Artificial Intelligence (AI) in the global response to the COVID-19 pandemic. With diverse detection methods in use, AI, particularly in the analysis of medical images, emerges as a critical component. Beyond its role in diagnostics, AI contributes significantly to investigations spanning pharmaceutical, molecular, and epidemiological dimensions, facilitating a comprehensive understanding of the virus. AI's versatility is evident in its application to various detection methods, including the analysis of respiratory specimens, blood serum, and chest images of infected individuals. It plays a crucial role in tracking and assessing the state of COVID-19 at multiple levels. Its ability to efficiently process and analyze vast amounts of data accelerates investigative efforts, making it an invaluable asset in the fight against the pandemic. Notably, AI enhances treatment stability and decision-making processes by evolving practical procedures and offering insights derived from data analysis. This extends beyond diagnostics, with AI actively contributing to the development of effective cure regimens and supporting the timely creation of vaccines, both essential components in the global strategy to curb the spread of COVID-19. Within AI, subfields such as Machine Learning (ML) and Deep Learning (DL) have proven instrumental in diagnosing COVID-19 cases. These subfields utilize mathematical models applied to medical images, particularly Chest CT and X-ray images, for tracking and diagnosis. Despite their effectiveness, current methodologies face challenges, including diverse image formats and an unbalanced dataset. To address these limitations, this research proposes an automated prediction methodology incorporating both Machine Learning and Deep Learning approaches. By utilizing diverse mathematical models and conducting experiments on three distinct datasets from reputable repositories, the proposed methodology aims to overcome the challenges posed by current techniques. The validation process underscores the proposed approach's supremacy in terms of accuracy, presenting a promising avenue for refining and advancing AI applications in the ongoing battle against the COVID-19 pandemic.

Keywords: AI, Framework, AI-Based Framework, Covid-19, Covid-19 Detection, X-Ray Image.

COVID-19: AN IODT ENABLED SUSPECT DETECTION AND PATIENT MONITORING FRAMEWORK USING DATA FUSION TECHNIQUES

Tayyaba Ilyas and Dr. Danish Mehmood

ABSTRACT

Recent developments regarding Internet of Things (IoT), Artificial Intelligence (AI), and Machine Learning (ML) opened new horizons of healthcare opportunities. These technological advancements give strength to face upcoming healthcare challenges. One of such challenges is the advent of Covid-19 which has adverse effects beyond comprehensions. The ongoing pandemic has frightened human health worldwide. Precautionary measures are expected to be necessary for the world to combat the Covid-19 disease outbreak until an efficacious vaccine is produced. Therefore, utilizing the basic functionalities of IoT's, this work presents a real-time rule based Fuzzy Logic classifier for Covid-19 Detection (FLCD). The proposed framework deploys the IoT framework to collect real-time symptoms data from users for early identification of symptomatic as well as asymptomatic Covid-19 patient. Moreover, the proposed framework is also able to monitor the treatment response of infected people. FLCD constitutes three components: symptom data collection using wearable sensors, data fusion through Rule Based Fuzzy Logic classifier, and cloud infrastructure to store data with a possible verdict (normal, mild, serious, or critical). After extracting the relevant features, experiments with a real Covid-19 symptom dataset are conducted to ensure effective and accurate detection of Covid-19 cases. FLCD successfully acquired 95% accuracy, 94.73% precision, 93.35% recall, and showed a minimum error rate of 2.52%. The results shows that the proposed FLCD framework is effective in diagnosing the Covid-19 suspect and providing an accurate identification of potential Covid-19 cases in terms of accuracy, precision, recall, and error. Employing FLCD can potentially reduce the impact of contagious virus, as well as the mortality rate through early diagnosis of the cases.

IMAGE BASED DEEP FEATURE EXTRACTION FOR THE DETECTION AND PREDICTION OF COVID19 THROUGH DEEP LEARNING

Arisa Rafiq and Dr. Muhammad Imran

ABSTRACT

The new strain of Influenza virus (SARS-COV-2) known as severe acute respiratory syndrome which is related to Covid19 has infected the entire international community. It has created severe health and safety concerns all over the globe. Identifying this form of the outbreak in the initial phase, may help in successful recovery with the advancement of computer analytics and particularly AI. It helps to relieve the moderated wellness program. The rapidly and exact identification of Covid19 limits the risk of spreading out of fatal disease. This research focuses on the deep feature extraction, accurate detection and prediction of Covid19. The proposed CNN concatenated model is based on Deep Learning models (Xception and ResNet101) for CXR images. For the extraction of features, CNN models (Xception and ResNet101) are utilized, and after that, these features combined by a concatenated model technique. In proposed scheme, the Particle Swarm Optimization method applied to the concatenated features that provides optimal features from the overall feature vector. The proposed scheme, also reduces the classification time. For the last classification, CNN is used. For the training of proposed model, we have used an enormous number of X-ray images. The dataset of CXR images was collected from three different sources. The results exhibit efficiency of proposed scheme for detecting COVID-19 with average accuracy 99.77%, 99.72%, and 99.73% on datasets 1, 2 and 3 respectively. Moreover, the proposed model also achieves COVID-19 average sensitivity 96.6%, 97.18% on dataset 2, and 98.88% sensitivity on dataset 3. Furthermore, 5-fold cross-validation provides better accuracy. The maximum overall accuracy of all classes including Normal, pneumonia, COVID-19 is about 98.02%.

Keywords: Image Based, Deep Feature Extraction, Covid-19 Detection, Covid-19 Prediction, Deep Learning.

EXPLOITING SHAPE AND MERGING OF DBSCAN ALGORITHM FOR THE IMPROVEMENT IN LARGE DATA

Sadiq Ullah and Dr. Sohail Asghar

ABSTRACT

DBSCAN is one of the most common technique of clustering in data mining for discovering groups of similar type data which is based on the density parameter. DBSCAN has ability to discover arbitrary and complex shape cluster and it is robust to noise in large spatial databases. Even with its advantages, DBSCAN algorithm has some problems such as this algorithm is unable to find the cluster with dissimilar density, it is also trouble in defining proper values for input parameter mean that total depends on input parameter. If the cluster are too close this algorithm is unable to distinguish proper clusters and DBSCAN cannot handle large size database his complexity in worst case is O(n2). Different variants of DBSCAN are proposed to overcome the above challenges because it introduced since 1996. Some researchers try to overcome the dependency on input parameter and propose reverse nearest neighbor same time, other tries to handle the complex objects with different structure & combination of DSets algorithm, some proposed sampling technique to handle large size data sets. The findings discovered this research area is still undeveloped and present literature absences the comprehensive review of current DBSCAN variations. The contribution of this study is to comprehensive analyze different DBSCAN variation & identify areas where further improvement is required to balance the performance of current algorithms. In this, study we also relate different variations with each other to choose the best algorithm and propose a new variant to overcome the drawback of Grid based DBSCAN.

Keywords: DBSCAN, Spatial Data, Clustering, Data Mining.

A PREDICTIVE ANALYTIC FRAMEWORK FOR LEVERAGING SMART HEALTH CARE

Ali Fida and Dr. Muhammad Usman

ABSTRACT

Accurate and timely prediction of the vaccine-preventable diseases is always a significant public health matter to reduce child mortality and morbidity rate. Pakistan has nation-wide different programs for timely treatment of vaccine-preventable diseases but unfortunately, coverage is quite low in spite of the accessibility of free vaccination and it influenced both maternal mortality ratio (MMR) and infant mortality rate (IMR). It's crucial for decision-makers to design effective strategies timely for reduction in MMR and IMR rates. A number of prediction models have been proposed in the past, however, the previously proposed models suffer from a number of limitations which place strong barriers toward the direct applicability of such models for accurate and timely prediction on large size healthcare datasets. In terms of prediction model building, firstly, the healthcare programs datasets have never been integrated in past to know the variables influencing on MMR and IMR across various healthcare programs. Similarly, the validation in past has been done on these programs separately, which do not give true representation of vaccine-preventable diseases. Secondly, the selection of features has been mostly done with the help of statistical methods only. The statistical methods have been successfully applied in different domains, however, these methods alone without the augmentation of domain knowledge have the tendency to yield erroneous results. Thirdly, the decision-makers have never been provided with an automated and visually enriched interface to timely predict the MMR and IMR based upon certain attributes selected before model building for accurate prediction. In order to address the aforementioned limitations, an automated and accurate prediction methodology has been proposed in this thesis, which overcome the said limitations in an effective way. The proposed methodology has been validated through an integrated approach by applying different machine learning algorithms. The proposed methodology provides an automated solution to decision-makers for timely detection of vaccine-preventable diseases to enhance the MMR and IMR rate along with the Maternal death and Infant Modality Rate. Experimental results have been conducted separately for MMR and IMR on different live datasets taken from Sindh healthcare programs and supremacy of proposed methodology has been observed by accurate prediction.

Keywords: Predictive Analytic Framework, Leveraging Smart Health Care.

A PREDICTIVE ACCEPTABILITY MODEL EFFECTIVE VACCINE DELIVERY IN EPI PROGRAM

Urooj Iqrar and Dr. Muhammad Bilal Qureshi

ABSTRACT

The immunization program is one of the most cost-effective and significant public interventions to vaccinate every child in the world and reduce child mortality and morbidity. The program is launched in Pakistan to protect infants and mothers against diseases and increase immunization coverage. Unfortunately, immunization coverage in Pakistan is still low. Certain factors contribute towards low immunization coverage, including lower acceptability of vaccines. Researchers from a range of different fields are carried out several studies to identify and analyze the reasons for low vaccine acceptability. Still, utilization of immunization data to improve coverage is rare. Several predictive analytic models were proposed to improve immunization coverage, mainly focused on vaccine hesitancy, vaccination uptake and defaulters identification. This study -proposes a highly interpretable predictive acceptability model to predict the acceptance of the vaccine accurately. Primarily, the contribution of this study is to classify vaccine acceptability into Low, Medium, Partial high and high to reinforce interventions by predicting acceptability for new and existing vaccines. Secondly, the study implemented the feature Importance method to make the model highly interpretable for Healthcare providers and vaccinations. Thirdly, this study finds the association of several demographic and socio- demographic characteristics with vaccine acceptability to help policymakers take initiatives to improve immunization coverage. This research has applied several Machine Learning algorithms, namely, Decision tree, Random Forest, Light GBM and XGboost, to Pakistan Demographic and health survey dataset (2017-18) and achieved 98% accuracy.

Keywords: Childhood immunization, Vaccination, Vaccine acceptance, Machine Learning, Feature Importance, Association Rule Mining.

A SMART FRAMEWORK FOR PHISHING URL DETECTION

Sajjad Jalil and Dr. Muhammad Usman

ABSTRACT

Phishing is a persistent and major threat on the internet that is growing steadily and dangerously. It is a type of cyber-attack, in which phisher mimics a legitimate website page to harvest victim's sensitive information, such as usernames, emails, passwords and bank or credit card details. To prevent such attacks, several phishing detection techniques have been proposed such as AI based, 3rd party, heuristic and content based. However, these approaches suffer from a number of limitations that needs to be addressed in order to detect phishing URLs. Firstly, features extracted in the past are extensive with a limitation that it takes a considerable amount of time to extract such features. Secondly, several approaches select important features using statistical methods, while some propose their own features. Although both methods have been implemented successfully in various approaches, however, these methods produce incorrect results without amplification of domain knowledge. Thirdly, most of the literature has used pre-classified and smaller datasets, which fail to produce exact efficiency and precision on large and real world datasets. Fourthly, the previous proposed approaches lack in advanced evaluation measures. Hence, in this thesis, effective machine learning and deep learning framework are proposed. Both approaches have gained popularity in detection of phishing URLs. However, the recent methods do not provide any information about the feature selection criteria and appropriate hyper-parameter settings for phishing URL detection that can enhance the efficiency of the model. Thus, in this thesis, URL only method is proposed, which is independent of any 3rd party services and utilizes only the URL for feature extraction. Moreover, the study aims to compare eight state of the art machine learning classifiers and six state of the art deep learning models for the detection of phishing URLs. The experiments are carried out on several benchmark datasets and compared the performance of machine learning classifiers against deep learning models to output the best model for detection of phishing URLs. The comparative results show that the proposed framework has achieved better precision and accuracy for phishing URL detection.

CLASSIFICATION OF THE HEALTHY AND UNHEALTHY BRAIN THAT SUFFERS FROM MCS AND UWS PATTERN DURING SLEEP USING MACHINE LEARNING TECHNIQUES

Muhammad Saad and Dr. Muhammad Imran

ABSTRACT

Determining the level of consciousness in patients with disorders of consciousness (DOC) presents significant challenges due to the complex patterns of neuropathology and variability between individuals. Resting-state fMRI (rs-fMRI) and EEG have been widely used to detect activity variations in DOC patients, but the identification of reliable diagnostic markers for consciousness level remains a work in progress. Misdiagnosis rates between unconscious wakefulness syndrome (UWS) and minimally conscious state (MCS) are notable, highlighting the need for improved assessment methods. Machine learning techniques offer promise in distinguishing between different states of consciousness, and this study explores their effectiveness in classifying UWS and MCS conditions. However, the use of EEG and fMRI data poses challenges in data collection, management, feature extraction, and dealing with class imbalance, high dimensionality, small dataset size, and noisy data. Ensemble learning methods, which combine multiple classifiers, have shown promise in addressing these challenges and improving classification accuracy. This dissertation focuses on utilizing ensemble learning techniques to overcome the complexities of biomedical datasets. The proposed ensemble machine learning methods outperform existing state-of-the-art techniques, achieving higher accuracy rates (94%, 91%, 95%) in classifying DOC states. These results demonstrate the potential of machine learning approaches for enhancing diagnostic capabilities in assessing consciousness levels in patients with disorders of consciousness.

Keywords: Healthy Brain, Unhealthy Brain MCS Pattern, UWS Pattern, Sleep, Machine Learning Techniques.

TEXT MINING FOR ABUSIVE CONTENT DETECTION: A TWITTER CASE STUDY

Syed Adeel Mushtaq and Dr. Fadia Shah

ABSTRACT

Due to rapidly growing interest of people on social networks, people subscribed to these social networks communicate and share their thoughts with other subscribers to a great extent. The most popular social networks, Facebook, Twitter, Instagram, and LinkedIn, etc, have billions of users of various cultures and backgrounds. It is necessary to identify the abusive content, so the subscribers feel safer while using these platforms positively. The content on these social sites is in text, images, audio, and videos in multiple formats. Audio, images, and video content are beyond the scope of this study. Social network Twitter is selected in this study due to the popularity and easy access to data. The data on Twitter is in an unstructured form so extracting useful information from unstructured data is usually not easy. Therefore, we are investigating the different existing text mining approaches to detect abusive content from online social networks. The study aims to propose a method to detect abusive content in the Twitter data that contain tweets, retweet, and comments. The prime objective of the study is to propose a new model to detect abusive content in Twitter content through feature extraction and also measure the level of abusiveness. The uniqueness of the study is that it identify the position of the abusive words which is missing in the existing studies because of various examples found where the position of the abusive word changes the context of the sentences. Abusive language is categorized based on features through the feature extraction process and categories either abusive or not. Three popular machine learning classifiers SVM, Naïve Bayes are LSTM used for classification to organize the abusive content into six categories; aggression, hate, cyberbullying, harassment, vulgar, and general. The general class is used to identify the content that is not abusive; remaining five classes are subtypes of abuse. The Evaluation of classifiers measures with accuracy. This research also provides a comparison with exiting studies to see evaluate the efficacy and accuracy of existing approaches in detecting abusive content. The purpose is to highlight the pros and cons of each approach that will be helpful to take care of important steps while developing new frameworks or methods to detect abusive content effectively in the user content on social networks. Secondly, this comparison is also helpful in figuring out the strategy most appropriate for determining the level of offensiveness to improve accuracy.

Keywords: Machine Learning, Abusive Languages, Twitter Social Network, Text Mining, Feature Extraction, Sentiment Analysis.

FACIAL BASED GENDER CLASSIFICATION FOR REAL TIME APPLICATION

Anmol Haider and Dr. Muhammad Imran

ABSTRACT

Appearance and facial features play an important role in gender recognition through images. For gender classification, multiple techniques were presented to acquire better results in which preprocessing part is one of the major and very important for gender classification as it removes noise, enhances, images, and eliminates any unnatural colors from an image. Another major aspect is the efficient feature extraction method. If features extracted accurately then the result of classification will improve. Over the past few years, gender classification techniques work perfectly for a controlled environment. However, challenges occurred for real-time applications due to low resolution, off-angle poses, faces with occlusion, and various expressions. The main focus of this study is to overcome existing challenges and propose a method that can be implemented in real-time applications. In this research novel method in which CNN has been used for classification of gender for real-time application as well as static images and video-based gender classification. An experiment has been performed in three databases Groups, GVLC and LWF. This research achieved 98% of accuracy.

Keywords: Gender Classification, Limitations, Recognition, Feature Extraction, Preprocessing, Real-Time Application.

UNDERWATER IMAGE ENHANCEMENT USING MODIFIED CLAHE AND ADAPTIVE COLOR CORRECTION METHOD

Ahmad Fraz Ashraf and Dr. Muhammad Imran

ABSTRACT

Oceans provide us with a lot of vital resources from food to energy, transport to defense. However, we still haven't utilize the full capacity of the field and it has much more to offer and only way to utilize it to capture visible images and videos but images and videos obtained from underwater are degraded. So it is important to enhance images and videos before processing. Enhancement is the way to improve or increase image quality and improve contrast of degraded images. Enhancement is one of the most researched and problematic area in image processing and most importantly underwater images are a very challenging zone for the photography field specially with its so many applications such as study about fishes, coral reefs, to capture 3D bathymetry of seafloor terrain and finding aquatic plants or saving them from extinction etc., The original image or video which is captured through image processing devices needs to improve in terms of intensity levels but there are various challenges such as less light available, low resolution and blurriness caused by the normal camera, these things create blurriness, noise, low contrast, color diminishing and less visible image. Various researchers have proposed their solution to overcome these problems. However, there are still some problems exists that needs to be solved. This report included reviews of the most used underwater image enhancement techniques and also introduces a method to overcome limitations in current literature. Dark channel Prior (DCP) is one of the most commonly used technique due to better Peak Signal to Noise Ratio (PSNR) value. However, DCP have some issues such as it tends to darken images, reduced contrast and produce halo effects. This method solve these issues by stretching contrast then applying Contrast Limited Adaptive Histogram Equalization (CLAHE). To validate the proposed method results will be compared with well-known state of the art techniques. This research viably addresses the image enhancement issue of the image processing systems. Comparing different methods made it validate the better.

Keywords: Enhancement, Color Diminishing, Contrast, Fusion Principle, Color Balancing technique, Contrast Limited Adaptive Histogram Equalization, Dark channel Prior, White Balance algorithm, Gamma Corrected Algorithm, Sharpened Algorithm, Wavelet Fusion algorithm, Mat-lab GUI Program.

IMPROVING SOFTWARE ARCHITECTURE EVALUATION

Tahira Sarfraz and Dr. Arshad Ali Shahid

ABSTRACT

Architecture of the software is described as a very important part of software development, since it is an asset for organizations that helps developers to build a high-level design for the software system. It prescribes the design decisions to fulfill user requirements that are to be implemented in the software that is being developed. The architecture of the system delivers quality attributes, so the final decision taken in the architecture design stage has a larger impact on the software system. Software architecture evaluation reveals that nonfunctional requirements (quality attribute) are being met or not. The software architecture is required to be evaluated at an early stage (before design and development) in order to make sure that the system is promising the required quality attributes. There are a number of architecture evaluation methods (e.g. SAAM, ATAM, ALMA, ARID, SBAR, ISAAMCR, and SAAMCS), we selected for software architecture evaluation are SAAM and ATAM. In this study, literature review is carried out to join to find out gaps and limitations of the existing methods/techniques SAAM and ATAM, and an improvement is proposed. The proposed improvement is "Improving Software Architecture Evaluation". The improved method consists of some steps of SAAM and ATAM along with one improved step that is Architecture Spike in it. The proposed improvements aim at covering common quality attributes including runtime and non-runtime quality attributes, creating a prototype, and identifying the potential risks and trade-offs in the architecture.

Keywords: Software Architecture, Architecture Evaluation techniques, SAAM (scenario-based architecture analysis), ATAM (trade-off analysis method), and Architecture Level Modifiability Analysis (ALMA).

IMPROVING THE SOFTWARE REQUIREMENTS TRACEABILITY FRAMEWORK ACROSS THE PRODUCT LIFECYCLE

Amna Ikram and Dr. Arshad Ali Shahid

ABSTRACT

Software Development Life Cycle (SDLC) can be defined as a process of software industry which designs, develops and tests software of high quality. The goal of this process is to generate a good quality software according to client needs and reaches completion within times and cost estimates. Requirements management (RM) is referred as the procedure of collecting, analyzing, refining, and prioritizing requirements of software product and planning for delivery of product. The main purpose of RM is to confirm that the company validates and meets the requirements of its clients plus external and internal stakeholders. It depends on the relationships among requirements and other software specifications. During management, requirements traceability is a key area to work in. Software projects deal with the rapidly changing objectives, goals and requirements of the client. Requirements traceability defines and follow the changes made in requirements in both forward and backward directions. It enables bottom-up tracing and top-down tracing. The main purpose of this research is to promote and explore the possibility of improving the traceability mechanism. However, different techniques of requirements traceability have been proposed and are being used to meet the changes required by different stakeholders. The support for the requirements traceability reduces the effort of development (hence cost and time) of impact analysis and also improves the change implementation. In this study, we have proposed a Lifecycle based Requirements Traceability (LRT) framework. LRT can link artifacts of all phases (Requirement, Design. Code, Test Case) of SDLC. It used the keyword based information retrieval method for the traceability of artifacts. It's a useful framework for the requirements traceability.

Keywords: Software Development Life Cycle, Requirements Traceability, Web Ontology Language, Pre-Requirements Specification, Post-Requirements Specification, Identification And Realization of Trace Relationships Within Requirements, Unified Modeling Language, Information Retrieval, Latent Semantic Indexing, Vector Space Model, Software Configuration Management, Software Change Control Board, Request for Change, Lifecycle Based Requirements Traceability.

A PREDICTIVE MODEL FOR HIV/AIDS BASED ON SOCIO-DEMOGRAPHICAL, BIOLOGICAL AND BEHAVIORAL DATA USING DATA MINING TECHNIQUES

Sehar Un Nisa and Dr. Azhar Mahmood

ABSTRACT

Acquired Immune Deficiency Syndrome and Human Immunodeficiency Virus (HIV/AIDS) remains to be key global health issue at top priority. HIV status knowledge aids both individual as well as community. Data collection of HIV is one of the most challenging factor associated to this disease. Despite the availability of several surveillance centers in various location of Pakistan, people are unaware of their HIV status. As social stigma associated with this disease, we cannot obtain accurate data from affected individuals. Unlike other disease, we are not supposed to relate any medical factors with this such as tumor detection, heart attack associated with heart disease. If a person has a risk behavior such as injecting drugs and sexual practices then there is a maximum chance of him/her getting infected with HIV/AIDS. Because of these factors, estimation of HIV/AIDS and prediction hold great importance. This work compares the traditional statistical approaches and data mining techniques along with prevalence estimation of a community with HIV status of individual entity. These estimations are based on few focused dimensions such as demographic, sexual behavior, injecting and other biological dimension. Among several possible solutions, one of them is to use statistical model to estimate HIV disease and identify risk factors in community.   These estimations are built on top of statistical models such as Logistic Regression etc. to approximate HIV prevalence, incidence and diagnosed/undiagnosed cases. Each model has their deficiencies which leave a margin of improvement for other statistical analysts to estimate prevalence trends. Another solution is to use data mining techniques so as to predict population's HIV status, which will help to detect the part of population which is affected the most to support and aid HIV prevention programs. Major Data Mining techniques such as Neural Network, Random Forest, J48, Decision Tree and Support Vector Machine (SVM) etc. are discussed and compared with their contributions and induced limitations for target problem. This study particularly focus on the multiple regions of Pakistan for HIV disease burden. The objective of this study is to conduct comparative analysis of previous methods and techniques focusing on HIV/AIDS prediction and estimation using statistical and knowledge discovery. It also presents a predictive HIV-model to scale up the acquired knowledge extracted of HIV status with 80% accuracy having an improvement of 10-15% as compared to dominant classifiers i.e. SVM, Neural Network, J48 and PART. HIV-model also generate an error score to validate predictive model and prevalence results to ensure HIV and prevent society from catastrophic spread of this epidemic.

Keywords: HIV/AIDS, Demography, Sexual behavior, Injecting, Prevalence, Predict, Data Mining.

MULTI-STAGE BREAST CANCER PREDICATION MODEL USING SUPERVISED CLASSIFICATION TECHNIQUES

Huma Qayyum and Dr. Azhar Mahmood

ABSTRACT

Breast cancer is a primary causes of death among the women worldwide. In 2017, an estimated 252,710 new cases of invasive breast cancer will be diagnosed among women early predication of breast cancer not only cure the cancer but also control the death rate. Breast cancer predication is difficult due to incompatible, inconsistence and noisy datasets. The main objective of this research to identify the crucial attributes which helps to identify Multi Stage of Breast Cancer. There is many attribute which helps in predication of breast cancer within datasets but there is a need to identify those attribute who helps in breast cancer stage predication. There is need to get better result in a case of breast cancer stages predication. There are several classification technique such as Support vector machine (SVM), Multilayer Perceptron and deep belief Network and many others techniques used for predication. These techniques are working effectively but the accuracy compromise under the following conditions; variability in datasets, multi-collinearity issue and sharp boundary problem. A lot of Method had proposed for a breast cancer predication but multi stage predication model need to be proposed. In proposed predictive Model we predict breast cancer stages by using available attribute of Wisconsin dataset. Breast tumor has two type benign and malignant caner benign has not cancer cell but malignant has cancer cell which spread to other part other part of body and it's difficult to recover Detection of cancer at the benign stage and prevention from spreading to other parts in malignant stage could save a person life's.

Keywords: Breast Cancer Detection, Preprocessing Techniques, Malignant Tumor, Benign Tumor, PCA.

EPILEPTIC SEIZURE DETECTION USING MACHINE LEARNING TECHNIQUES

Ayesha and Dr. Muhammad Bilal Qureshi

ABSTRACT

Epilepsy is a global neurological disorder characterized by abnormal electrical activities in the brain, resulting in epileptic seizures. Electroencephalography (EEG) is commonly employed to detect seizures, but manual analysis is time-consuming and requires continuous expert supervision. To address these challenges, this research proposes an automatic epileptic seizure detection model using machine learning techniques. The model utilizes feature extraction methods to capture temporal, spectral, non-linear, and pattern features from EEG signals. A feature selection strategy is employed to identify the most distinguishing and significant features. Classification into normal, interictal, and ictal classes is achieved through various traditional and fuzzy logic-based classifiers. The proposed model is evaluated using benchmark datasets, demonstrating significantly improved accuracy compared to existing methods. For single-channel EEG recordings, RF and FRNN achieve the highest accuracy of 98.74% and 98.89%, respectively, for interictal versus ictal classification. For multi-channel EEG signals, DT and FURIA algorithms yield the best performance with 99.85% and 99.92% accuracy, respectively. The validation results confirm the consistent performance of the selected classifiers on unseen test datasets, providing reliable epileptic seizure detection.

Keywords: Epilepsy, EEG, Machine Learning, Seizure Detection, Epileptic Seizure.

SDN BASED DDOS PROTECTION IN SMART GRID SYSTEM

Sh Hassan Mahmood and Dr. Danish Mehmood

ABSTRACT

Traditional power grid system has evolved recently into Smart Grid (SG) after emergence of Internet of Things (IoT). SG is now faced with both classical and evolving cyber threats due to involved diverse and complex heterogeneous Information and Communication Technology (ICT) environment. Advanced Metering Infrastructure (AMI), constituting significant portion of utility-consumer network, is most vulnerable to network attacks like Distributed Denial of Service (DDoS) attack. Existing security technologies, besides their efficiency, are not adequate enough to cater for DDoS security in SG due to highly distributed, heterogonous, and dynamic network. Detection mechanisms have to rely on localized available data, which is a major bottleneck in achieving DDoS protection in distributed networks like SG. Same is evident from the recent attack on Ukraine power grid. Recently, emerging Software Defined Networking (SDN) based approaches for DDoS protection have been proposed by researchers to cater for bottlenecks in traditional network paradigms, but still they are providing protection against flooding attacks only and are dependent on static thresholds. Proposed SDN-based DDoS Protection System (S-DPS) is efficiently addressing these issues by employing a light-weight entropy based defense mechanism using SDN environment and adaptive threshold for SG. It provides not only early detection mechanism within 250 packets, but mitigation of both Low-Rate and High-Rate DDoS attacks in real-time as an addition, achieving detection rate of 100% with false positive rate of 0% for simulated scenarios. Moreover, minimal impact on controller's CPU / RAM utilization has been observed due lower computational complexity of approach used.

Keywords: SG, SDN, OpenFlow, Entropy, DDoS Attack.

A DATA-DRIVEN FRAMEWORK FOR INTRODUCING PREDICTIVE ANALYTICS INTO EXPANDED PROGRAM ON IMMUNIZATION, PAKISTAN

Sadaf Qazi and Dr. Muhammad Usman

ABSTRACT

Immunization is a significant public health intervention to reduce child mortality and morbidity. Pakistan has a nation-wide Expanded Program on Immunization but the coverage is quite low in spite of the accessibility of free vaccination. A number of studies have been done to identify the reasons for this low coverage but analysis and utilization of immunization data to increase the coverage is rare. A predictive analytics model has been proposed in the past to improve the coverage by identifying the children who are most likely to miss any of the vaccines included in the immunization schedule, known as defaulters, but it suffers from a number of limitations. Firstly, there was no categorization of high and low coverage areas on the basis of defaulters for prioritized vaccination. Secondly, it only classified children into two stages: defaulters and non-defaulters, considering all children at high risk of defaulting even if he misses only one dose. Thirdly, there was no association between defaulters and their demographic and background characteristics like region, place of delivery, number of children under 5 in a household and wealth index of the family, to better understand the reasons behind defaulting. Motivated by the aforementioned limitations, this study proposes a defaulter prediction framework for the accurate identification of defaulters. Primarily, our contribution is to classify defaulters at five different stages: defaulter, partially-high, partially-medium, partially-low, and non-defaulters to reinforce targeted interventions by accurately predicting children at high risk of defaulting from immunization schedule. Secondarily, our model categorizes high coverage and low coverage areas on the basis of a number of defaulters for more targeted interventions. Furthermore, this study finds associations between defaulters and their socio-economic and demographic characteristics to help decision-makers in designing strategies for increasing coverage by reducing the number of defaulters. Different machine learning algorithms are applied to Pakistan Demographic and Health Survey (2017-18) dataset and achieved 95% accuracy.

Keywords: Machine Learning, Childhood Immunization, Predictive Analytics, Vaccinations, Multilayer Perceptron, Association Rule Mining.

EFFICIENT DATA STREAM MINING WITH CLASS IMBALANCE AND CONCEPT DRIFT

Affan Ahmad Toor and Dr. Muhammad Usman

ABSTRACT

With the continuously growing usage of the internet and smart devices, more and more data streams are being generated, and stream mining is becoming a bigger field. Among other challenges faced by data streams, a major challenge is concept drift, which is caused by an overtime change in data distribution. Detecting concept drift becomes trickier when the class distributions in data are skewed. This phenomenon is known as class imbalance where the distribution of the class labels is highly uneven. Data streams often have both these problems, but previous works have rarely acknowledged and handled them together. This study provides a detailed review of efforts done to solve these problems, both in isolation and in the presence of each other. In this study, a systematic approach is adopted to come up with a strategy to handle concept drift with class imbalance in data streams. This study proposes a framework that is capable of handling both class imbalance and concept drift for data streams. The proposed framework, Enhanced Reactive Drift Detection Method (ERDDM), takes a sliding window from the incoming data stream and removes the class imbalance from the instances using SMOTE techniques. Then the resultant data is passed on to the drift detector, which calculates the error rates to determine whether there is any kind of drift available in the data. If the drift had already started in leading windows, then it is checked if the drift is ended or still continuing. If there is no current drift, then a new drift is initialized, and its history is maintained. The proposed drift detector also infers the type of drift, i.e., Abrupt or Gradual. The detector then passes this information to the classifier to build a model based on the drift type. After the classification is done, the results are verified based on measures like prediction error, detection delay, total detected drifts, and mean evaluation time. Two separate experimentations are performed on synthetic and real-world datasets. For synthetic datasets, four datasets generators are utilized, which are most common in existing literature. Six different sizes of these four datasets are generated to check the scalability of the proposed method on different stream sizes. These experiments were repeated for data with class imbalance and without class imbalance. The results showed that ERDDM has outperformed its predecessors and showed the least prediction error. Similar trends are seen for real-world datasets as well. Call Detail Record (CDR) data of a telecom company is used for the experiment on a real-world dataset. It is a known fact that the churn prediction data naturally contains class imbalance, and according to domain experts, concept drift also exists in CDR data. This combination makes the CDR data a perfect fit to be used in this study. According to our knowledge, such an experiment has never been done on CDR data for data stream mining. The novelty of this experiment is itself a significance of this study. We believe that using this study CDR data can be utilized in a more meaningful way, and it can also be used for further studies in the stream mining domain.

Keywords: Efficient, Data Stream Mining, Class Imbalance, Concept Drift.

AN EFFECTIVE MODEL FOR MULTI-LABEL LEARNING IN WEB CATEGORIZATION

Malik Irfan Shokat and Dr. Muhammad Usman

ABSTRACT

This thesis deals with the subject of Multi-label learning in Web Categorization. Multi-label learning comes under the broad umbrella of Multi Relational Data Mining techniques. In multi-label learning, each instance is linked with multiple labels and the aim is to predict the most suitable label set for the unseen instance. Multi-label learning is used by modern applications, such as text classification, functional genomics, image classification, web classification, music categorization, in this work, special interest has been given to the area of multi-label learning in web classification. Web page classification, also known as web page categorization, is a specific type of text categorization, where the content of each web page is utilized, to classify it into a set of predefined categories. The categorization of web page content is very important for many information retrieval tasks. Since web pages have text and multimedia data, they can be viewed as structured, semi-structured or unstructured. This has imposed additional challenges to web page classification as compared to traditional text classification. In this thesis, we have reviewed approaches that have been recently used, for web categorization to solve the multi-label problem. Most of the authors have only used two evaluation metrics: Precision and Accuracy. In our research, we did not find anybody who had used scalability as an evaluation metric. Past researchers have performed experiments on benchmark and small datasets, but these do not provide a true representation of a real world web database. Also, we tried to find a tool/application that could automatically classify web pages into multiple categories or labels concurrently, but no such tool/application exists. To overcome the above-mentioned limitations, we proposed an effective model for multi-label learning in web page categorization using the Natural Language Processing (NLP) techniques. We performed a series of experiments on twelve benchmark datasets and two real world multi-label learning websites using our proposed model. To evaluate the performance of our model for web categorization, we used five well known multi-label classifiers, as well as using various evaluation metrics like Precision, Accuracy, and Hamming Loss. Our experiments on the benchmark datasets and real world enterprise applications showed that the classifier Pruned Sets with a Threshdd (PST) is the best multilabel classifier for web categorization. A web-based application has been developed on the basis of the proposed model for multi-label learning in web categorization. This application can categorize any real world website's pages into multiple categories or labels. The application has a user-friendly GUI to provide fast and accurate results, which not only save time but also save the manual cost of data labeling.

Keywords: Multi-Label Learning, Web Categorization, Multi Relational Data Mining, GUI, PST.

AUTOMATED TEST CASE GENERATION OF SMARTPHONE APPLICATIONS USING MODEL BASED APPROACH

Usman Habib Khan and Dr. Muhammad Naeem Khan

ABSTRACT

Automation of testing process not only reduces testing time and efforts of applications testers but also improves correctness and accuracy of the testing process. Mobile applications have their own quirks regarding testing, such as the high number of different events that need to be tested. Though several steps in the overall testing process had been automated but still there is a room of improvement and novelty. Due to an evolving domain, it carries its own challenges that should be kept in mind while planning to test smartphone apps. As smartphone apps moderate in size so model based testing using state machines and state charts could be a promising option for maximum coverage and maximum completeness of test cases. Using model-based testing approach, we can automate the tedious and very important step of generating test cases which ultimately results in minimum errors, maximum coverage and completeness of test cases in addition to saving time for the overall testing process. In this study we explore application in targeted fashion and model application's most critical modules which help avert generating repeating test cases, thus improving efficiency in terms of time and cost. We also checked the effectiveness and impact of using model-based testing approach in the domain of smartphone applications. Test cases have been generated on a real device of the targeted model of the application under test and results are compared with other contemporary testing approaches.

Keywords: Android Application Testing, Model Based Testing, Functional Testing, Smartphone App Testing, Test Case Generation.

A FUZZY BASED APPROACH TO CONSTRUCT INTELLIGIBLE CHURN PREDICTION AND RETENTION MODEL FOR TELECOM CUSTOMERS

Muhammad Azeem and Dr. Muhammad Usman

ABSTRACT

Accurate and timely identification of the potential churner, also known as churn prediction is crucial to devise effective retention strategies. A number of churn prediction models have been proposed in the past, however, the previously proposed models suffer from a number of limitations which place strong barriers toward the direct applicability of such models for accurate prediction on real world large size telecom datasets. In terms of prediction, firstly, the feature selection methods adopted in majority of the past work neglected the information rich variables present in Call Detail Records (CDRs) for model development. Secondly, the validation on existing models is performed on benchmark or small datasets which do not provide a true representation of a real world telecom datasets. Thirdly, the selection of features has been mostly done with the help of statistical methods only. The statistical methods have been successfully applied in different domains, however, these methods alone without the augmentation of domain knowledge have the tendency to yield erroneous results. Finally9 the previous studies has completely neglected the Fuzzy classification methods which perform reasonably well in terms of accuracy for large datasets. Moreover there is very limited amount of work reported in literature that has extended the prediction models towards automatic and intelligent retention mechanism. In addition, the complaints data is never integrated with potential churners to devise an effective retention strategy. The categorization and severity of the predicted churners has not been focused in the past for targeted and intelligent retention campaigns. In order to address the aforementioned limitations, a prediction and retention methodology has been proposed in this thesis which overcome these limitations in an intelligent way. The proposed methodology has been validated using large size C'DR datasets of a South Asian Telecom company. The proposed methodology generates intelligent retention campaigns by mining customer usage and complaints patterns. Experimental results on two real world telecom datasets revealed the supremacy of proposed methodology and managed to retain 87% of the potential churners.

Keywords: Fuzzy Based Approach, Intelligible Churn Prediction, Retention Model, Telecom Customers.

MULTI-ASPECT BASED OPINION CLASSIFICATION MODEL FOR TOURIST REVIEWS

Muhammad Afzaal and Dr. Muhammad Usman

ABSTRACT

In the era of digital connectivity, tourists worldwide rely on social media platforms such as Google, Facebook, TripAdvisor, and Expedia to gather information and opinions about tourist destinations. Traditional methods of seeking recommendations from friends and family have been replaced by online reviews. Opinion mining techniques play a crucial role in classifying these reviews as positive or negative. However, considering the diverse aspects associated with each tourist place, there is a need for opinion mining methods that can identify and rate various aspects mentioned in the reviews. This has led to the emergence of aspect-based opinion mining as a new area in social reviews mining. These methods not only identify the aspects referred to in reviews but also classify opinions about these aspects into different polarity classes. While aspect-based opinion mining has proven useful in domains such as products and movies, the tourism domain still faces several challenges. Firstly, aspect extraction requires domain knowledge and manual effort. Secondly, implicit and co-referential aspects further complicate the extraction process. Lastly, the classification of multi-aspect opinions, where multiple aspects are discussed in a single review with varying sentiments, poses a significant hurdle. This thesis proposes an effective aspect-based opinion mining model that addresses these challenges. To automate the aspect extraction process, an extension is introduced that utilizes aspect-sentiment based patterns to extract explicit aspects and eliminate irrelevant ones. Implicit aspects are handled using a hierarchy-based method that associates commonly used sentiment words with the appropriate aspect in a hierarchical structure. Co-referential aspects are identified and grouped using a probabilistic co-occurrence based method that calculates aspect similarity. Furthermore, a multi-label learning approach is proposed to tackle multi-aspect opinion classification, accurately classifying opinions into polarity classes. The effectiveness of the model is evaluated through experiments on benchmark and real-world datasets, employing various multi-label classifiers. The results demonstrate the superiority of the proposed model, achieving 90% accuracy per label in multi-aspect opinion classification and correctly extracting 87% of aspects. Additionally, a visually enriched and user-friendly prototype application is developed to showcase the effectiveness of the proposed model.

Keywords: Multi-Aspect Based Opinion, Classification Model, Tourist Reviews.

MULTI-LEVEL PATTERN EXTRACTION AND PREDICTION IN LARGE DATASETS

Muhammad Usman and Dr. Muhammad Usman

ABSTRACT

This research thesis addresses an underexplored area in the field of knowledge discovery by combining data mining and data warehousing techniques. While these domains have been independently utilized for knowledge discovery, recent efforts have focused on hybrid approaches, which offer advantages such as reduced reliance on expert guidance and the ability to mine at multiple levels of abstraction. Additionally, extracted patterns are evaluated for their interestingness and can be visualized. However, little work has been done in an integrated fashion, and existing techniques for pattern extraction and prediction have several limitations. The selection of important features in the mining process is inefficient, manual processing at multiple levels of abstraction is cumbersome, advanced evaluation measures applicable to multi-dimensional environments are lacking, and effective visualization support for large patterns is lacking. To address these challenges, a novel methodology is proposed in this research thesis. The methodology allows for the extraction and prediction of patterns at multiple levels of abstraction, handling both numeric and nominal data in a unified manner. It comprises two algorithms for automating multidimensional schema generation and data loading, eliminating the need for manual processing at multiple levels. The extracted and predicted patterns are evaluated using advanced measures, and a visualization component enables interactive analysis of the patterns. Experimental studies conducted on real-world datasets demonstrate that the inclusion of different variable ranking techniques improves pattern quality compared to similar approaches. The automation of multidimensional schema generation also enhances efficiency, and the visualization component enables effective analysis of a large number of extracted patterns. These findings highlight the effectiveness of the proposed methodology in extracting informative and diverse patterns, predicting patterns at multiple levels of abstraction, and supporting interactive analysis.

ADVANCED A-JUMP WITH GRID COMPATIBILITY

Adeel-ur-Rehman and Dr. Naveed Riaz Ansari

ABSTRACT

Architecture for Java Universal Message Passing (A- JUMP) is a Java based message passing implementation. A-JUMP offers flexibility for programmers in order to write parallel applications making use of multiple programming languages. There is also a provision to use various network protocols for message communication. A High Performance Computing (HPC) bus acts as the backbone of A-JUMP. The HPC bus serves to facilitate development of parallel application and has been built upon Java Messaging Service (JMS) using pure Java programming language; which are well-established standards of software development industry nowadays. ActiveMQ (an implementation of JMS by Apache) is exclusively in charge for message passing and data communication in an asynchronous fashion. Certain data manipulation as well as data communication tests had been carried out in order to measure the performance of the framework as compared with some of its predecessor models like MPICH2, MPJ Express etc. The results for standard benchmarks like Ping-Pong latency, Embarrassingly Parallel (EP) code execution9 JGF Crypt etc. gave us the conclusion that for the cases where the data size is smaller than 256KB. The numbers are comparative with the other compared models. But, in case, the packet size exceeds 256K bytes, its performance seems to be severely hampered. So, considering this observation, AA-JUMP (Advanced A- JUMP) emphasizes on improving the performance of the original A_JUMP. The second aim of AA- JUMP is to extend its scalability to the computational Grid scope so that the jobs/processes that were having inter-process communication on cluster or multi-cluster level may be able to function in a real grid environment. The milestone is achieved by testing it over the Worldwide LHC Computing Grid (WLCG) setup and basic results are obtained.

Keywords: A-Jump, Grid Compatibility, High Performance Computing, MPICH2, MPJ Express.

OPTIMIZED FEATURES SELECTION FOR GENDER CLASSIFICATION USING COMPUTATIONALLY INTELLIGENT TECHNIQUES

Sajid Ali Khan and Muhammad Nazir

ABSTRACT

Gender classification is an application of classification that has received a great deal of attention recently. However, the techniques presented so far in different literature dealing with gender classification still have a significant room for improvement. It is an easy job for human to recognize the gender but challenging task for machine. One small image contains thousands of pixels there for to extract the important facial information some intelligent techniques are required. To select the best features extraction technique, we analyzed different techniques. Goal of the thesis is to propose an efficient frame work for gender classification which is robust to different facial variation like facial expression, image contrast and occlusion. It has been concluded after performed analysis with different techniques that the performance of various techniques changes with change in dataset. In-order to minimize the error rate. Different classification techniques has been combined. Large data size can make the technique slower and less accurate to produce optimal results therefore Particle Swarm Optimization (PSC) is used to provide more discriminative features sets and eliminate redundant features. Principle Component Analysis (PCA) has also been utilized for dimension reduction. With the use of these well-known powerful techniques computational complexity has been reduced with enhanced accuracy Experiments are performed using Stanford University Medical Student (SLJMS) and 1.1 W face database. This thesis covers essentially the research work on the following: 1) Literature review of Gender classification techniques using face gait and body-based databases, 2) Efficient feature selection mechanism has been developed for gender classification, 3) Different experiments on FERET databases in-order to reduce the data dimensions. , Optimized features selection using Particle Swarm Optimization techniques, 4) Enhancing the classification accuracy rate after classifiers ensemble, 5) Development of framework for gender classification using real world face images.

Keywords: Optimized Features, Gender Classification, Computationally Intelligent Techniques.

EFFICIENT DATA HIDING TECHNIQUES WITH ENHANCED CAPACITY

Mehdi Hussain and Dr. Mureed Hussain

ABSTRACT

Digital communication has become more popular due to tremendous growth of internet. The digital communication has many advantages as it has brought new challenges and new opportunities for innovations. Information security is one of the most significant challenges. Information is exchanged between parties within few seconds. General information is transferred by different applications such as web browsing, streaming of audio, videos, bank transaction software or etc. That's why security of such information is very important for that application or communication aspects and it has been a major concern nowadays. Mostly countermeasures of information security are achieved through encryption and data holding techniques. This thesis discusses the available data hiding methods for digital information with respect to images and network communication protocols. It addresses both theoretical and practical aspects of information hiding methods for both types of digital media, and also proposes new data hiding methods. In first part, we briefly discuss the literature review of existing best available method of data hiding with respect to communication and image domains. Various embedding methods target different aspects like perceptual transparency, capacity and robustness or etc. Our focused is based on two major aspects: transparency and capacity aspects based on available methods. We have proposed different comprehensive solutions to achieve data hiding in images and as well as network communication protocol. For targeting perceptual transparency, we have proposed an edge base data embedding with high (peak signal to noise ration) PSNR method, u-here steno-image (carrying hidden data) has very high resemblance to cover-image original image). For high capacity aspect, we have proposed pixel intensity based high capacity data embedding method. In communication protocol context, we have explored the utilization of the packet length and also the packet payload to achieve high rate of hidden data. The experimental results of proposed methods confirm our theoretical calculations and high capacity of data hiding in our proposed techniques.

HIDING SENSITIVE SML ASSOCIATION RULES USING BAYESIAN NETWORK

Khalid Iqbal and Dr. Sohail Asghar

ABSTRACT

The increasing electronic storage of information and the need for standardized data sources have led to the emergence of the Xtensible Markup Language (XML) as a global standard for information exchange and presentation. However, standardization brings the risk of disclosing sensitive information, which can undermine the value of shared data from a business perspective. Privacy Preserving Data Mining (PPDM) has garnered attention as a means to hide sensitive information disclosed by Association Rules, enabling the rapid and efficient discovery of knowledge without compromising privacy. Existing rule mining processes rely on assumed sensitive items and do not leverage Bayesian Networks (BN) in the XML domain. This thesis proposes a PPDM model that preserves XML Association Rules (XARs) by utilizing Bayesian Networks for probabilistic-based privacy preservation. The model identifies sensitive items based on the probabilistic occurrence and frequency in the original transactional data source, enabling the transformation of the data source into a modified version. The modified data source can then be used for information sharing with the collaboration of the apriori algorithm, which investigates the dataset horizontally to generate XARs. The proposed model ensures reliable identification of sensitive items, eliminating the need for assumptions. It provides a mathematically robust approach to determine item sensitivity without solely relying on support and confidence measures. The significance of this model lies in its ability to automatically identify and protect sensitive items, thereby preserving privacy while generating XARs for collaboration.

Keywords: Sensitive SML, Association Rules Using, Bayesian Network.

AUTOMATED SERVICES DISCOVERY FOR INDUSTRIAL APPLICATIONS

Aakif Nazeer Khan and Dr. Sohail Asghar

ABSTRACT

Emanti has emerged as a new paradigm for meaningful data retrieval in distributed environments, and Semantic Web Services (SWS) have automated information processing, addressing limitations in current technologies. The telecommunications industry, with its increasing service demands and dynamic nature, requires service providers to support service automation for cost reduction and improved response time. This thesis reviews various service discovery techniques and identifies limitations in existing approaches. The advent of SWS has reduced manual work and human involvement in service discovery, composition, and selection. Industrial surveys and interviews with Pakistan-based telecom operators and service providers highlight existing challenges faced by the industry. The objective of this thesis is to overcome industry limitations and provide an integrated enhancement in semantic web services, spanning the entire service lifecycle. The proposed model integrates goal-based service discovery, semi-automatic service composition, and mediation. Web Service Modelling Ontology (WSMO) and domain-specific ontologies are adopted, along with the development of a Software as a Service (SaaS) based billing application prototype for dynamic service discovery in a distributed environment. Validation experiments using telecommunications data from Pakistan demonstrate the advancement achieved by integrating Web Service Execution Environment (WSMX) and Internet Reasoning Service (IRS-III) in the services lifecycle. Evaluation results show significant reductions in time, cost, and human effort. Strengths, weaknesses, opportunities, and threats (SWOT) analysis highlights the benefits of implementing semantic and SaaS-based applications in the telecommunications industry. The proposed integrated services discovery model surpasses prior architectures, although it acknowledges the complexity of addressing all types of heterogeneous services with similar functionalities across different providers. The thesis work sets forth a comprehensive approach but recognizes the limitations in achieving the ultimate service discovery model within its scope.

Keywords: Automated Services, Discovery, Industrial Applications.

AUTOMATION OF THE HOSPITAL MANAGEMENT SYSTEM OF KALSOOM NURSING HOME

Safi Ur Rehman and Muhammad Nadeem Khokhar

ABSTRACT

The project is about constructing a paperless nursing home. The idea behind is that most of the companies or nursing homes of today communicate and keep their record stored on paper, The main disadvantages of which is that it occupies small space, communication is slow and this source are unreliable, inefficient and insecure. Therefore an alternative system is required which can cover these drawbacks. A complete organization is imagined which is not depend on paper for communicate and storage. This software supports the department of Kalsoom Nursing Home, as I have to automate the entire process. The software is theoretical, feasible economically and technically. This software intends to give a new feel to the existing nursing home through comprehensive software. The main aim of software is to provide them software that can help them perform their function quickly and easily with optimum performance. The development by has been progressed using structural, objected oriented approach. The software has been implemented using SQL Server 2000 as back end, Visual studio Do Net 2005 and Seagate Crystal Report for reports.

Keywords: Automation, Hospital Management System, Kalsoom Nursing Home.

OPTIMAL RESOURCE SELECTION IN GRID ENVIRONMENT

Muhammad Aqeel and Dr. Mansoor Alam Ansari

ABSTRACT

The ultimate objective behind the construction of any Grid is to better utilize the existing available resources for the jobs which require huge amount of processing power and generate tremendous amount of data to be stored. Selection of appropriate resources for executing the user jobs is one of the most important features of any Grid environment because non-optimal resource selection may spoil the ultimate objective of the Grid environment. The objective of this thesis is to study the resource selection mechanism in Grid computing and to propose a new matchmaking mechanism which can select resources nearer to user requirements. In order to select appropriate resources, user describes its request for resources using some resource specification language in existing technologies. These specifications lead to a lot of ambiguities in resource description which may lead to selection of undesired resources. The resources with attribute values too higher than the required values can be selected with existing matchmaking mechanisms. This is not optimal and may leads to wastage of resources. The proposed mechanism helps the Grid users to select the most appropriate resources according to user specified criteria and avoids the wastage of precious computing resources like imputing cycles, storage (both primary and secondary) etc. In existing matchmaking mechanisms, all the resource attributes are treated equally whether they have great impact on better execution of user job or not; which may results in selection of sources which do not completely match the user specified criteria. The proposed system solves this problem as it has suggested a resource ranking mechanism that gives a clear picture of the resource Optimal Resource Selection in Grid Environment. As thousands of computers can be connected to a grid at the same time, it is possible that a peat number of computers may satisfy the user request. In this case, the new matchmaking mechanism selects the resources based on their reliability factor. This reliability value of any resource is calculated by applying fuzzy rules on the values provided by fuzzy sets of past performance, usage of the resource and fulfillment of the contract with the user. Existing matchmaking mechanisms use different resource specification languages which are not portable across different platforms and thus makes it difficult to redirect the user request as one cluster to another if the request resource is not available in the cluster whose resource specification language is used. In order to solve this problem, new matchmaking mechanism uses XML for both resource description as well as resource request from the user. XML based resource specification language is platform independent and can be used easily across heterogeneous environments.