Signals and Communication Technology

Pradeep Kumar Singh ·
Marcello Trovati · Fionn Murtagh ·
Mohammed Atiquzzaman ·
Mohsen Farid *Editors*

Data Science and Artificial Intelligence for Digital Healthcare

Communications Technologies for Epidemic Models

Springer

Signals and Communication Technology

Series Editors

Emre Celebi, Department of Computer Science, University of Central Arkansas, Conway, AR, USA

Jingdong Chen, Northwestern Polytechnical University, Xi'an, China

E. S. Gopi, Department of Electronics and Communication Engineering, National Institute of Technology, Tiruchirappalli, Tamil Nadu, India

Amy Neustein, Linguistic Technology Systems, Fort Lee, NJ, USA

Antonio Liotta, University of Bolzano, Bolzano, Italy

Mario Di Mauro, University of Salerno, Salerno, Italy

This series is devoted to fundamentals and applications of modern methods of signal processing and cutting-edge communication technologies. The main topics are information and signal theory, acoustical signal processing, image processing and multimedia systems, mobile and wireless communications, and computer and communication networks. Volumes in the series address researchers in academia and industrial R&D departments. The series is application-oriented. The level of presentation of each individual volume, however, depends on the subject and can range from practical to scientific.

Indexing: All books in "Signals and Communication Technology" are indexed by Scopus and zbMATH

For general information about this book series, comments or suggestions, please contact Mary James at mary.james@springer.com or Ramesh Nath Premnath at ramesh.premnath@springer.com.

Pradeep Kumar Singh • Marcello Trovati •
Fionn Murtagh • Mohammed Atiquzzaman •
Mohsen Farid
Editors

Data Science and Artificial Intelligence for Digital Healthcare

Communications Technologies for Epidemic Models

Springer

Editors
Pradeep Kumar Singh
Department of Computer Science
and Engineering
Central University of Jammu
Samba, J&K, India

Fionn Murtagh
Department of Computer Science
University of Huddersfield
Huddersfield, UK

Mohsen Farid
University of Derby
Derby, UK

Marcello Trovati
Department of Computer Science
Edge Hill University
Ormskirk, UK

Mohammed Atiquzzaman
School of Computer Science
University of Oklahoma
Norman, OK, USA

ISSN 1860-4862 ISSN 1860-4870 (electronic)
Signals and Communication Technology
ISBN 978-3-031-56817-6 ISBN 978-3-031-56818-3 (eBook)
https://doi.org/10.1007/978-3-031-56818-3

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Switzerland AG 2024

This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

If disposing of this product, please recycle the paper.

Preface

Introduction

The digital revolution in healthcare is a significant area gaining importance worldwide due to technology advancements. AI-enabled medical devices, diagnostics and treatment, telemedicine, and block chain-based electronic health records have led to a digital revolution in healthcare. In particular, mobile digital healthcare is another crucial area which is gaining considerable popularity with significant potential to further innovate digital healthcare, as well as provide an important contribution to patients' wellbeing. There are many practical examples where practitioners utilize digital health technologies such as iECG, handheld ultrasound, lab on a chip technologies, clinical diagnostics and treatment.

Distributed healthcare cloud is a recent and fast-growing area of development in digital healthcare, due to its cost efficiency, high scalability, availability and support to new ways of developing and delivering services. Furthermore, it is expected that edge computing will actively contribute to a further digital healthcare enhancement. However, these examples are by no means exhaustive as new research, technology and integrated frameworks are further pushing the current research frontier. For example, 5G mobile networks and edge computing may become a substantial enabler of the future personalized digital healthcare ecosystem, whilst artificial intelligence approaches will enable the introduction of nurse robots, chatbots and virtual health assistants. Moreover, AI is used in precision medicine, medical imaging, drug discovery, genomics and for predicting illness and diseases. The large quantity of data, which is continuously created, provides a useful tool for preventive care, accurate staffing and to medication or clinical error prevention, by identifying, extracting actionable information and insights. Wearable technologies are another dimension used for collecting patient health data from the medical devices, including wearable technologies for improving health services.

This book may be useful to the scientists, practitioners, scholars and researchers who are working in the field of computer science, engineering and communication engineering, along with the students in these subjects who are working or willing to work on Data Science and Artificial Intelligence for Digital Healthcare. Specialists as well as student readers will find the book chapters encouraging and helpful. The book is organized into four major parts: (i) Pandemic Models Using Data Science and Artificial Intelligence; (ii) Artificial Intelligence and Machine Learning In Healthcare 4.0; (iii) IoT, Edge/Fog and Cloud In Digital Healthcare; and (iv) Distributed Ledger, Security Solutions and Innovative Models for Digital Healthcare.

The Objective of This Book

The prime objective of the proposed book is to explore the current areas of research and development, and challenges in the area of digital healthcare using recent technologies such as data science and artificial intelligence. The main objective of this book is to identify the high quality, original, technical contributions in the field of mobile digital healthcare using recent technologies: AI, deep learning, IoT and distributed cloud computing. Digital Healthcare has become a crucial research area due to the availability of enormous quantity of structured and unstructured data. Furthermore, the ongoing COVID-19 pandemic has further demonstrated that the digitalization of health-related information and resources can make a significant positive impact.

Therefore, new technologies and methods are being developed to further enhance relevant applications and theoretical frameworks. This book will address a significant proportion of such advances to enable a prompt information sharing across researchers and practitioners. In particular, mobile and distributed healthcare cloud technologies are emerging trends with enormous potential to generate significant innovative solutions. This book will provide an in-depth analysis of such aspects. Several practical examples will be discussed, including (but not limited to) iECG, handheld ultrasound, lab on a chip technologies, clinical diagnostics and treatments.

Distributed healthcare cloud is a recent and fast-growing area of development in digital healthcare, as it offers low cost, high scalability and availability and supports new ways of developing and delivering services. It is expected that edge computing is going to transform digital healthcare service up to next level, and there are endless possibilities to research further. 5G mobile networks and edge computing may become a substantial enabler of the future personalized digital healthcare ecosystem.

This book is to provide a unique resource to the readers who are seeking to understand and explore key technologies on digital healthcare. This book is useful to understand the digital healthcare methods and technologies. It provides few of the examples and case studies that are focusing on new areas to fully engage the readers who wish to explore new findings in the area of data science, artificial intelligence and its usages in digital healthcare world. This book may be a useful resource as a reference for M.S. (Masters) and Ph.D. scholars for exploring new dimension of research in the area of digital healthcare.

Organization of the Book

There are 16 chapters in this book, which are divided into 4 parts. (i) Pandemic Models Using Data Science and Artificial Intelligence; (ii) Artificial Intelligence and Machine Learning In Healthcare 4.0; (iii) IoT, Edge/Fog and Cloud in Digital Healthcare; and (iv) Distributed Ledger, Security Solutions and Innovative Models for Digital Healthcare.

In the first part of the book, first chapter focuses on "Post COVID-19 Remote Medicine and Telemedicine Evaluation via Natural Language Processing Techniques".

The second part of the book covers the chapters related to Artificial Intelligence and Machine Learning In Healthcare 4.0 and the topics include (i) Leveraging Computer Vision for Gender, Age, and Ethnicity Prediction Using Deep CNN; (ii) Swarm Intelligence: A Segmentation Approach; (iii) Smart Management Based on Deep Data Analysis for Digital Healthcare; (iv) An Intelligent Diagnosis System Based on SVM with Dragonfly Metaheuristic Algorithm for Preventing and Predicting Hepatitis C Infection; (v) An Automatic Diagnosis System Based on Machine Learning Models for Predicting Hepatitis C from Blood Samples; (vi) Digital Healthcare System Using Stacked Ensemble Machine Learning Model to Predict Heart Diseases; and (vii) Emergence of Bayesian Network as Data Imputation Technique in Clinical Trials.

The third part of the book covers the chapters related to IoT, Edge/Fog and Cloud in Digital Healthcare, and topics under this part are (i) Smart Implementation of IoT and UAVs-Based Transportation of Blood Samples for Digital Healthcare; (ii) AI and UAVs in Smart Transportation of Urgent Organ Transplant and Usage for Organ Donor and Donee; (iii) Real-Time Organ Status Tracking System for Digital Healthcare; and (iv) IoT and AI-Based Smart Healthcare Monitoring System

Finally, the fourth part of the book contains the chapters which focuses on Distributed Ledger, Security Solutions and Innovative Models for Digital Healthcare, and the chapters are based on following topics: (i) Exploring the Impact of a 5E-Flipped Learning Environment on Students' Learning Motivation: A Case Study of Medical Assistant Education; (ii) Building Interoperable Electronic Health Records as Purpose Driven Knowledge Graphs; (iii) Mobile Digital Solution for Road Safety Through ECG Analysis of Driver's Anxiety; and (iv) The Intervention of Artificial Intelligence in the Healthcare Sector: Trends and Challenges.

This book explores the capabilities of data science and artificial intelligence technologies and its impact on many unique and innovative solutions for the digital healthcare. The ideas, concepts, methodologies, algorithms discussed in this book may found to be useful for the researchers, scientist, healthcare professionals and could be utilized for making better digital healthcare services worldwide.

Samba, J&K, India — Pradeep Kumar Singh
Ormskirk, UK — Marcello Trovati
Huddersfield, UK — Fionn Murtagh
Norman, OK, USA — Mohammed Atiquzzaman
Derby, UK — Mohsen Farid

Contents

Part I
Pandemic Models Using Data Science and Artificial Intelligence

Chapter 1
Post COVID-19 Remote Medicine and Telemedicine Evaluation via Natural Language Processing Techniques

Marcello Trovati, Tariq Soussan, Yannis Korkontzelos, and Nikolaos Polatidis

1.1 Introduction

According to Institute of Medicine, delivery of healthcare can be improved by following six specific goals, which are *safe, effective, patient-centred, timely, efficient, and equitable* [1]. Realising the importance of information technology, in September 1999, the committee on the quality of US healthcare identified five areas which can enhance the delivery of healthcare performance and its above identified goals, namely *access to medical knowledge-based, computer-aided decision support system, collection and sharing of clinical information, reduction in errors, and enhanced patient and clinician communication* [1]. Recently the COVID-19 pandemic has significantly highlighted the delivery of healthcare services. In particular, many hospitals were overwhelmed with patients infected with coronavirus infections and hospitals had to close normal delivery of face-to-face physical consultations for out-patients to protect its staff and patient both from this infection. Hospitals and clinics resorted to delivery of basic healthcare services by offering them through remote consultations or telemedicine.

M. Trovati (✉) · Y. Korkontzelos
Department of Computer Science, Edge Hill University, Ormskirk, UK
e-mail: trovatim@edgehill.ac.uk; Korkonty@edgehill.ac.uk

T. Soussan
Department of Information Technology College of Computer Engineering and Science, Prince Mohammad Bin Fahd University, Dhahran, Saudi Arabia
e-mail: tsoussan@pmu.edu.sa

N. Polatidis
School of Architecture Technology and Engineering, University of Brighton, Brighton, UK
e-mail: N.Polatidis@brighton.ac.uk

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024

P. K. Singh et al. (eds.), *Data Science and Artificial Intelligence for Digital Healthcare*, Signals and Communication Technology,
https://doi.org/10.1007/978-3-031-56818-3_1

WHO defines telemedicine as the *"use of information and communication technologies to improve the patient outcomes by increasing access to care and medical information"* [2]. By employing this technology, access, quality, equity, and cost-effectiveness should be addressed to the satisfaction of the patient. Telemedicine has been identified as a technology which uses the information and communication systems to provide easy access of healthcare facilities to patients, thus improving the delivery and reach of these facilities to every community [3]. Furthermore, it also includes the use of online real time video calls to diagnose, treat, advice, follow-up and prescribe a patient on their medical conditions [4]. It gives a new meaning to providing basic healthcare as one does not have to travel to clinic nor wait for longer times to get a face-to-face appointment with a doctor. It can also be used for diagnosis of basic medical conditions, follow up on treatment and monitoring of any medical condition progress, via already built digital platforms such as Microsoft, Skype, Whatsapp, or specific dedicated platforms can be designed to cater for special needs of the patients and practitioners [4]. However patients need to have access to the internet to use telemedicine and be willing to meet their doctors in this way.

In order to optimise the opportunities of this technology, there are several companies which have offered solutions to healthcare providers by providing devices which can be remotely used in places such as care homes, patient's home and children hospitals. They use multiple hand held devices which carries out *"clinical level physical exams of the heart, lungs, skin, ears, throat and abdomen, and measures body temperature and heart rate, to enable remote diagnosis of acute care situations"* [5]. The results are captured by these devices, either video or audio, which are then communicated back to consultants or specialists. These devices can be operated by nurses at care homes, patient homes, remote locations or even by patient themselves.

Most of the information captured during clinical examinations are in the form of textual hand written notes either on paper or punched into computers in the form of patient's electronic health records, previous health history, diagnosis notes, medication prescriptions, and lab results. This type of information is classified as unstructured data which is not in the form of traditional structured data of tabular rows and columns found in databases and worksheets [6]. Since language used in medicine is highly specialised, it is necessary to extract the right words, combination of words and conceptual relations which can then be fed to the machine learning (ML) and deep learning (DL) models to aid in the decision making for clinicians through use of such telemedicine devices.

Natural language processing (NLP) is a research area, which focuses on the extraction of concepts, and their combinations, and mutual semantic and syntactic relations from textual unstructured data. In this chapter, PubMed database and Twitter were used to search and extract relevant insights by employing various NLP techniques.

This chapter is structured as follows: Sect. 1.2 gives an introduction on the main challenges related to telemedicine, with particular emphasis on text mining techniques, which are further discussed in Sect. 1.3. Specific information extraction techniques are presented in Sect. 1.4 and the datasets used for extraction and results obtained are discussed and evaluated in Sects. 1.5, 1.6. Finally Sect. 1.7 concludes the chapter and prompts to new research directions.

1.2 Related Work

Telemedicine refers to remote consultations (RC) appointments which takes place between patient and healthcare professional either through telephone or via internet [7]. Such consultation take place via video appointments via suitable platforms, such as Skype, WhatsApp, FaceTime, and MS Teams. The onus of RC appointment is on healthcare professionals to ensure it is secure and if required, patients have the appropriate support available during the appointment including family members witnessing the call, and the technology is fully working. The aim is to enable the patients to feel relaxed and comfortable during this type of consultations.

The General Medical Council (GMC) has identified factors which can determine when RC is appropriate or not [8]. There are three principles which need to be followed: *"Standard of good practice should apply to both face-to-face and RC, if standards of good practice cannot be applied to RC, then shift to face-to-face, and lastly agreement must be reached with patient for choosing over the resources available for consultation"* [8]. Consent and continuity of care have been identified as key issues by GMC for advice and treatment via RC.

Telemedicine applications are usually designed and implemented by using mobile phone and web-based platform designed on PHP and My SQL, where images would be taken via mobile technology and sent to the appropriate specialist for further analysis [9]. This has helped to provide access to screening in remote areas where specialists healthcare providers were not readily available for face-to-face visits. A study about patient's and clinician perception with the delivery of telemedicine services was conducted in 2012 for one year in a US hospital setting among 15 clinical departments in which around 62% patients and 59% clinician reported satisfaction with the service [10]. The study was conducted among the established patients who belonged to remote or underserved areas. Patients viewed these appointments to save travel time, appointments scheduling and were even willing to pay for this type of service due to remoteness of their location or hinderance due to patient's physical ability in travelling to the hospital. However, clinicians also pointed out the technological issues related to connectivity during administering of such type of service in some cases.

Similarly in Australia, a seven-week study was conducted for telemedicine nurses to make them better understand the importance of self-reflection and teach-back after every session with a patient using telemedicine facility [11]. This facilitated the understanding of the use of better communication skills to compensate for the lack of face-to-face interaction. After every shift, health practitioners had to self-reflect on what went well, how could they have improved the experience and what was the patient perception of use of such facility. The teach-back session was then implemented in the form of video analysis, discussions, and role play. This improved the overall satisfaction among the nurses to further improve their interaction with the patients through this medium.

A four-questionnaire based empirical study was conducted over five years for patients, who were fitted with pacemakers, as follow-up treatment in Spain [12]. The aim of the study was to find out the patient's perception of the experiences and communications during the telemedicine monitoring versus the conventional way of face-to-face monitoring. The patients were divided into two groups to find out their experiences and communication perceived over telemedicine monitoring vs conventional face-to-face monitoring. Both groups generally reported positive experiences and telemedicine monitoring group reported better treatment diagnosis which was being suited to their needs. However, both groups were of the view that they wanted more autonomy in their treatment, less waiting time, and better management by doctors.

Patient activation, which refers to patient knowledge skills and confidence in managing own health condition, and self-efficacy, which refers to one's ability to performs certain behaviours to reach specific goals, were measured for patients diagnosed with Inflammatory Bowel Disease over a year by use of telemedicine [13]. The patient interaction was through a web-based platform with additional support of text messaging. However, the use of such facility did not appear to improve either the patient activation, or the self-efficacy as compared with the face-to-face interaction. Although telemedicine has been in use before the COVID-19 pandemic in different countries, setups, and departments, its importance has significantly appreciate during the pandemic.

For example, telemedicine visits to a healthcare network of 4 hospitals in US increased from daily 80 visits to more than 1000 visits in a space of 15 days [14]. The visits included both for the urgent and non-urgent care with less than 60% of the urgent and less than 20% of the non-urgent visits were related to COVID-19. The average age of the patients accessing these visits was between 20 and 44 years old. This increase was helped by the fact that the healthcare network had a same information and communication infrastructure for all the healthcare facilities under it by which patients' health records were easily accessible to any healthcare provider of this network. Plus, the capacity of healthcare staff administering these visits were also increased. The patients were able to easily access the healthcare app and pay for utilising a video call with a healthcare provider.

Remote Paediatrics-Gastroenterological clinics in US shifted to telemedicine practices once the restrictions of the COVID-19 pandemic were set [15]. These visits were offered by nurse practitioners (NP) who lacked experience in delivery of such service. After every such visits, feedback from NP were taken as well as feedback surveys were also conducted from the parents. The surveys were found to be consistent with the feedback from the NPs. Generally, parents complained about incomplete assessments, lacking face-to-face intimacy and technological issues related to connectivity during such visits but, overall, they were satisfied with such type of healthcare offering. A short telemedicine facility was administered for less than 3 weeks during COVID-19 for cardio-oncology patients in a clinical setting in which 11 patients including new and existing were seen by the healthcare clinicians [16]. All patients were offered this facility upon referral and during these appointments further physical appointments were also assessed based upon patient's condition and diagnosis in which case they were either referred to their local ERs or an in-person visits were arranged. During these appointments, due care was given to interpersonal skills while interacting and communicating with the patients. Moreover, it was also made sure that due to the nature of telemedicine consultation, healthcare clinicians were licensed to practice at the location of the delivery of the service which was being offered by them.

Telemedicine can also be used for the benefit of in-patients in a hospital setting. To protect the staff and in-patient from risk of infection during pandemic, a tertiary hospital in Israel divided its internal medicine unit and intensive care unit into two zones each classifying them into contaminated and clean zones [17]. The contaminated zones would contain in-patients infected with virus and staff in PPE whereas the clean zone would act as control room and infection free area where staff would support their colleagues in the contaminated zone through telemedicine facilities utilising both audio and visual technologies thus reducing the risk of transmission of virus.

Apart from pandemics, telemedicine has also been used in different specialised healthcare like palliative and addiction-control, and healthcare for senior people. A mobile app was designed, and training was provided to family and community care providers of patients requiring Palliative home-care [18]. The training about the app helped the family care providers to upload the data related to patient well-being on to the app which in turn was assessed by the Palliative care provider teams on a data dashboard. This helped in addressing the patient concerns, better utilisation of resources and more targeted feedback and training to family care providers in the usage of the app for better patient management. This also helped in the reverse feedback in the app improvement.

Telemedicine has also been associated with a positive impact on the decrease consumption of alcohol carried out from the review of studies which were carried out in US, EU, and Australia [19]. Apart from this, the use of such healthcare facilities has also reduced depression in the people who are struggling with the consumption of alcohol, increased their quality of life and increased patient satisfaction among other positive outcomes. The facilities used were telephone calls, video calls, mobile and web-based applications.

An empirical study was carried out in retirement homes and rehabilitation centres in the Ontario, Canada region to find out the trust perceived by the elderly people in using the machine learning governed healthcare systems and the study also revealed how this factor coupled with personal characteristics can increase people use of such healthcare facilities [20]. This survey-based study was carried out among elderly people from 3 retirement homes and 2 rehabilitation centres and were asked about their perception of healthcare services received through both face-to-face and through an machine learning (ML) based healthcare setup. The result from the study found that for such a ML based healthcare system to be successful, it must be easy to be learn and operated plus the user/patient needs to have a feeling of control, social interaction, and belongingness especially for elderly people residing in care homes. Similarly, elderly users of online health systems who are between the age of 65–90 are perceived to acquire knowledge without the input of professionals and are less confident in their own assessment of their health [21]. This study was carried out in Norway where senior users have different view about accessing healthcare facilities which is to see a doctor, only when need arises. Stress is pointed on introducing more accessibility of such applications for user of all ages and groups.

To fully exploit the benefits of telemedicine, it is also necessary to know about the barriers and hindrances which can create obstacles for true utilisation of such facility. A systematic review of literature has found that barriers to use of telemedicine varies from region to region however some common barriers are related to technological issues, technical proficiency of healthcare providers in handling such devices and facility, patient demographics such as age, education, availability and access to technology, costs, and change resistance [22]. Some measures have been suggested to overcome these barriers by targeted training and education of patient and healthcare providers and giving patient the option to choose between face-to-face and telemedicine interactions. Different drivers and barriers were identified by utilising a modified version of unified theory of acceptance and use of technology to better understand the application of telemedicine in India [23]. Drivers were recognised as *performance expectancy, effort expectancy, facilitating conditions, social influence, habits, hedonic motivation, and price value with government policy, top management, and project team capability as additional drivers* [21]. Barriers were identified as *financial, social, time, technology, and security and privacy risks* [23].

With the current advancement in field of computing mainly in the machine learning (ML) and deep learning (DL) models, this has opened doors to realisation of these specialised areas in healthcare. Though biomedical is a highly specialised field which utilises terms, abbreviations, literature, and concepts which are rarely used in commonly spoken or written free-text form. Therefore, the building of applications over ML and DL models requires extensive pre-processing of data associated with healthcare facilities like patient health records, clinician notes, test/lab results, medical history, and prescription diagnosis. For this purpose, text mining and more specifically, natural language processing (NLP) is being utilised to pre-process the unstructured data, in the form as described above, into a shape which consist of the right words or group of words along with the identified relations to the medical literature/datasets, and which can then be used in the ML/DL models for making the right judgement by the healthcare professionals [24, 25].

Text mining (TM) focuses on the extraction of useful information and the pattern associated with it from large data which can be found in the form of structured or unstructured [6]. Structured data consists of data which is found in the form of traditional rows and columns forming up a table like in databases or worksheets whereas unstructured data can be found in textual documents in any form, emails, and web. Text mining is used for research purposes where the goal is to intelligently extract key words from the text along with their pattern for their analysis and eventually building a knowledge base. Text mining has many applications in modern computing fields especially in NLP. NLP is a field in computing research that focuses on finding patterns and relations in the structured and unstructured data which can then help in making computer understands how human beings use language and then develop applications for targeted purposes. NLP has seen a growth in many areas like speech recognition, artificial intelligence, information extraction from medical reports and records, and language analysis [6].

NLP has been used to correct the spelling mistakes occurrence in the Electronic Health Records (EHR) documents. An NLP system was devised over two stages where the first stage has three sub-stages where the spelling of the word was checked by building relations and pattern and measuring the distance of the sample word with a created list of words [26]. After which the sample word would then be passed through next two sub-stages to make the interpretation of the word clear and ambiguous in terms of context and part of speech. Finally in the second stage the sample word would then be compared with the medical terminologies used by the clinicians to ascertain that the chosen and corrected word is the expected outcome. It was shown that by using such an NLP system improves the quality of the text in the EHR documents without any spelling mistakes.

As the biomedical literature is continuously growing, sometimes it is difficult to keep track on the terms and their context. This is where word or term identification can help in identifying the concerned term by relating it with the medical data sources and literature [27]. This is done in three steps, i.e., term recognition, term classification and term mapping. Term recognition is used to separate relevant word or a group of words from non-relevant terms recognised from the related literature and data sources. Term classification analyses these recognised terms according to their relevancy in the established medical categories and classes. This is necessary for the knowledge management and the correct mapping. Finally, term mapping allows the terms to be mapped with the medical databases and literature based on the identified relationships in the previous steps plus an identifier is also attached with it for reference. This completes the exact term identification process.

Since medication terminology and sentences are complex and are only associated with the field of medication therefore sometimes it is difficult for the NLP methods to correctly capture the words and their true relationships. By utilising two existing NLP systems, i.e., MLP (medical language processor) and MedLEE (Medical Language Encoding and Extraction System), the authors created a database of 4564 medication sentences from discharge documents and created a semantic based-probabilistic context free grammar model for building a structured dataset for these medication sentences which helped in finding the semantic relations between medical terms and this also helps in reducing any errors in understanding the final output [28].

NLP is being used to aid clinical decision support (CDS) systems which are used by the healthcare professionals in US to make decisions based on the recommendations generated from the patient data and knowledge-based resources [29]. Most of the time patient data is manually entered however the efficacy of such a system is increased when it connects with an EHR which can then help in raising alerts and warnings requiring actions from the healthcare providers. As most of the data is entered in an unstructured way which is present in the medical literature, lab reports, prescriptions, test results, medical history, admission/discharge notes, and clinician notes. NLP helps in analysing these data by looking for specific medical keywords and concepts to trigger warnings or recommendations for clinicians to decide further actions. Such a CDS system based on NLP can be active or passive, where the former will trigger warnings and alerts on its own based on the current data in the system, whereas the latter are activated only when specific information is entered like finding specific patient demographic in a community etc. Data used for both active and passive systems are in unstructured form as described above.

Word embeddings in NLP utilises the capture of words with similar vector representations thus showing the underlying pattern and relation between the connecting words like “cats and dogs, spoon and fork, tea and kettle”, etc. [30]. By using the word embedding method, medical terminology words captured from the patient EHR, and medical literature was compared with words from the web and news resources. It was found that the model trained on EHR, and medical literature was able to find more medical terms words which were acceptable by human beings as compared to words embedding trained from web and news resources.

Careful analysis of patient information, i.e., past medical history, tests, medication, and information provided on arrival at the Accident and Emergency departments (A&E) can be used by the NLP model to assess and predict whether patient will be requiring the usage of various advanced diagnostic imaging tests [31]. This will help to reduce over-attendance at the A&E, waiting time in analysing the medical information to ascertain the requirement of test and freeing up ADI resources to more critical needs. In this regard, US National Hospital Ambulatory Medical Care Survey data was used to determine the factors which are critical for deciding in the usage of ADI services [31]. Topic modelling and multivariable logistics regression models were used to extract the desired word and association of words based on medical terminology which were later used to measure their connectivity with the desired outcome predictors through the regression models. The model predicted the right outcome as desired and has the potential to reduce the wait times at the A&E.

With the advancement in deep learning models, it is becoming preferable to use them as it allows for information extraction and retrieval of related terms and concepts from clinical texts by using self-learning deep models. In this regard, four transformer models BERT, RoBERTa, ALBERT, and ELECTRA were tested over three public medical databases to carry out extraction of relevant terms and clinical concepts [32]. The result of these transformers was compared with an LSTM based model, and the study found out that transformer models perform better especially the RoBERTa in extracting the relevant terms and concepts from the clinical texts.

Another way by which deep learning models can be trained on extracting relevant information from clinical text is to use a generalised deep learning model which has been trained on extracting words or group of words from sources like Wikipedia, Google, or news [33]. Then transfer the learning of that model to a specialised deep learning model which will read the text from medical and clinical literature to extract the terms and concept which are specific to biomedical resources and data. This will save time in building special datasets or carrying out learning of a model on how to extract feature from such a dataset. However, question will remain on identification of specialised term recognition by the model since medical terms are highly specialised.

As evident from above literature that by employing NLP based pre-processing techniques, a system can be designed which can aid in the decision making of healthcare professionals and the five ICT related areas identified by the committee on the quality of healthcare in America can be made possible in an improved delivery of healthcare performance [34].

1.3 Text Mining Techniques Used in Healthcare

It is crucial to distinguish the data mining techniques that are most frequently utilised in the healthcare business because the tools of data mining do not all hold up the same data mining techniques [35]. In the below section, previous work on some of these data mining methods is discussed:

- *Classification*. Previous work has showed that two methods to classify health associated content. One is based on keywords while the other is based on learning. Some previous has been done on each method In the first method, a regression model of influenza-like illness can be assessed using the proportion of flu-related Google search queries over the same period [36]. Another work suggested a technique to correlate the number of flu-related tweets, discovered by flu-related keyword discovery, with the actual influenza-like-illness rates [37]. Other work suggested to utilize association mining and "the Proportional Reporting Ratios" the mine association between drugs and their averse reactions in social media. By identifying the existence of the healthcare keywords produced by utilising Consumer Health Vocabulary (CHV), they detected content having the adverse drug reactions [36]. Other approaches include converting identification problems into categorisation problems and using classifiers that are based on machine learning classifiers to categorise the data into classes. This approach works when it trains and learns with set of labelled texts, then uses the trained learner to categorise unlabelled texts [38]. Some previous work has been done as well on this. A previous work suggested an algorithm was proposed to detect tweets related to diseases, thus it filters tweets that include keywords that are defined as syndromic in the BioCaster public health ontology [39], followed by categorising the filtered messages into any of the predefined ailments through utilising the binary unigrams as features. Another work implemented a classifier based on Support Vector Machine to identify flu-related tweets [40]. It is taught with unigrams compiled within the same proximity of keywords related to flu.
- *Clustering*. Classification is defined as a data analysis technique assembling models that predict categorical labels [41]. Clustering techniques seek to group documents into clusters [42]. Topic modelling techniques aim to obtain topics

from a collection of text files based on statistical methods. Each theme is described as a distribution over a group of words [43]. Both clustering and topic modelling have features that are similar. The modelling of themes and clustering are methods among the suggested NLP techniques, used to infer patients' concerns, pursue new health-related reports, and discover emerging health themes [43]. Previous work has showed that applying clustering algorithms that utilise data from social media platforms have been done to biomedical research [44]. A previous work has showed that text clustering algorithms that use data from social networks has been implemented to determine health-related themes [42]. Another work has shown that it was applied to extract postings that are related to ADR [45]. Other work has shown that k-mean algorithms can be utilised for filtering ADR-related textual information from social media data [46].

- *Linear Regression*. Logistic regression is defined as a statistical method for analysing a dataset for a binary classification problem [47]. A relationship between a dependent binary variable and a minimum of one independent variable is determined with the help of linear regression [47]. Previous work has demonstrated that the logistic regression offered a better recall and F1 measure than support vector machine in classification tweets relevant and irrelevant to asthma [48]. Other previous has was related to the use of maximum entropy for the identification of tweets related to illness [49]. Another work utilised text data on autism spectrum disorder (ASD) obtained from Twitter in order to implement a "generalised linear regression models" (GLMs) to examine the impact of the wording of the text and the categorisation among various health subjects [50].

1.4 Information Extraction Via Text Mining Techniques

Due to the diverse nature of textual data, it is essential to provide information extraction capabilities from unstructured data. In this chapter, specific text mining techniques are considered to allow the identification and assessment of relevant information from textual data sources [51].

Depending of the general context and the given semantic information, a variety of text mining techniques can be used, which in general depend on the type of data and their structure. In particular, sentiment analysis [52], focuses on the detection of "opinions" or *polarity* from textual data sources. The method introduced in [53] specifically targeted a dataset containing information on air accidents and near misses [54]. More specifically, some of the entries consisted of pilots' comments.

In this chapter, we have expanded this method by, first of all, improving the vocabulary containing the keywords as in [53] and based on the techniques and methods discussed in [55–58]. These included a list of words suitably describing the associated polarity. An extensive set of new keywords and cue phrases was created by automatically extracting them from the tagged version of the Brown Corpus, which contains approximately 500 samples of English-language texts [59]. This was carried out by considering the triples `(NP1, VP, NP2)` where

Table 1.1 Examples of renormalisation as per [53]

Term		Normalisation
hesitated	→	hesitation
uneasy	→	uneasiness
don't know	→	uncertainty
be careful	→	lack of carefulness
be specific	→	lack of specificity

- `NP1` and `NP2` are the *noun phrases*, i.e., phrases with a noun as its head word [60], which had to contain one or more keywords from [53]. Note this requirement had to be satisfied for *at least one* of the `NP`s, and not just for both of them.
- `VB` is the *linking verb*.

Subsequently, the extracted `NP1` and `NP2` were manually analysed to identify the appropriate keywords, and cue phrases. A detailed evaluation of this approach goes beyond the scope of this paper, since it specifically addresses issues that are not directly relevant in this context. However, we tested it on two randomly chosen papers [61, 62]. The automatic extraction was compared with a manual one, which produced a recall of 65% and a precision of 74%.

Similarly to [53], the following steps were included:

- Textual fragments from input datasets were first shallow parsed via the Stanford Parser [51].
- A grammar-based extraction identified triples of the form `(NP, verb, keyword)`, where `NP` is the noun phrase, `verb` is the linking verb, and `keyword` consists of one or more keywords as mentioned above.

The triples are used to populate the nodes and edges of the corresponding network, by identifying any connection among the keywords defined above, with the corresponding elements of the datasets. In order to avoid any redundancy, all the extracted terms were normalised, where *normalisation* is the process of mapping different variants of a term to a unique and standardised form [60]. For example, Table 1.1 depicts instances of normalisation as described in [53].

In the above example, the term on the right hand side column would define connected individual nodes as part of the corresponding network.

In this paper, Improved Sentiment Urgency Emotion Detection (ISUED) model from previous work has been used [63]. This model is based on three classifiers to be joint into a single model which are sentiment analysis, urgency detection, and emotion classification [63]. This model is based on Multinomial Naive Bayes algorithm. The N-gram used for this model is Unigrams or words (n-gram size = 1) and Bigrams or terms compounded by two words (n-gram size = 2) [63]. The categories associated with the three classifiers are shown in Fig. 1.1:

The feedback categories can be from the customers which can be customer service reviews, product or service reviews, or proposals to enhance their service [63]. The same environmental experiment will be used as previous work, which is a

Fig. 1.1 Categories for the ISUED model [63]

Categories
Anger
Features
Complaint
Positive
Negative
Pricing
Feedback
Request
Happiness
Sadness
Hate
Surprise
Love
Worry

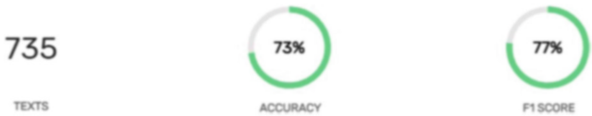

Fig. 1.2 Accuracy and F1 Score metrics for training the ISUED Model [63]

Fig. 1.3 Overall Keyword List from training the ISUED Model [63]

Windows 10 Enterprise laptop with a processor of Intel Core i5-8250U CPU. The installed memory (RAM) is 8.00 GB. The System type is 64- bit Operating System, x64-based processor [63]. The data size for a dataset is 347 recent Tweets where “GP consultations” has been mentioned up until 7th of November 2021.

This model from previous work has been trained to enhance accuracy and F1 score [63]. The accuracy is defined as the percentage of test tweets that were matched with the right category and it is the quotient of the correctly classified tweets by the overall tweets in test dataset. The F1 score combines both precision and recall [64, 65]. Recall is the proportion of positive sentiments which are correctly acknowledged while precision is the ratio between the correct sentiments predicted to the total number of matches predicted [64]. The training of the model from previous work resulted in Fig. 1.2 which is the accuracy and F1 score of the model, and this also resulted in Fig. 1.3 which are the keywords resulted from the training [63].

1.5 Description of the Datasets

The data used in this article comes for two main sources:

- PubMed [66], which is a free search engine mainly accessing the MEDLINE database of references and abstracts on life sciences and biomedical topics.
- Twitter.

The former was used to identify a large collection of abstracts based on the following keywords and their appropriate synonyms

- *Online patient monitoring*
- *Remote patient consultations*
- *Telemedicine*
- *Telehealth*
- *Impact of COVID-19 to remote consultations*

This identified over 50,000 abstracts, which were in a text file after any duplicated items were removed. Subsequent pre-processing was carried out, which included lemmatisation and removal of any stop word to ensure the textual data could be suitably further analysed.

Twitter data was obtained using similar terms, and 347 tweets were extracted. The aim was to complement the above textual data with information identified from social platform to analyse *how* these topics are discussed and shared.

1.6 Results

After the ISUED model's training was finished, the model was run on the dataset mentioned in previous section which contains 347 recent Tweets where the term(s) "GP consultations" have been mentioned. Processing the dataset shows the category or categories for each tweet and its/their respective confidence value(s) such that 25 different single or combined categories have been produced. Figure 1.4 shows the top 20 categories containing the largest group of tweets using the ISUED Model. Figure 1.5 displays the top 10 categories percentages out of the overall. It is shown that most of the tweets have been classified as "Feedback:Complaint" category. Some difficulties were noted. Some tweets were not able to be categorised correctly or were only partially categorised correctly. This could be due to false positive or false negative in the tweets or the model requiring more training.

For every tweet, the ISUED model computes the confidence value for each single category that matches it. Furthermore, the average confidence values of all the categories for each tweet is then calculated. Based on this, the average confidence value of each of the top 10 categories from Fig. 1.4 can be reached. This is defined as the average of the average confidences for all the tweets belonging to the same category.

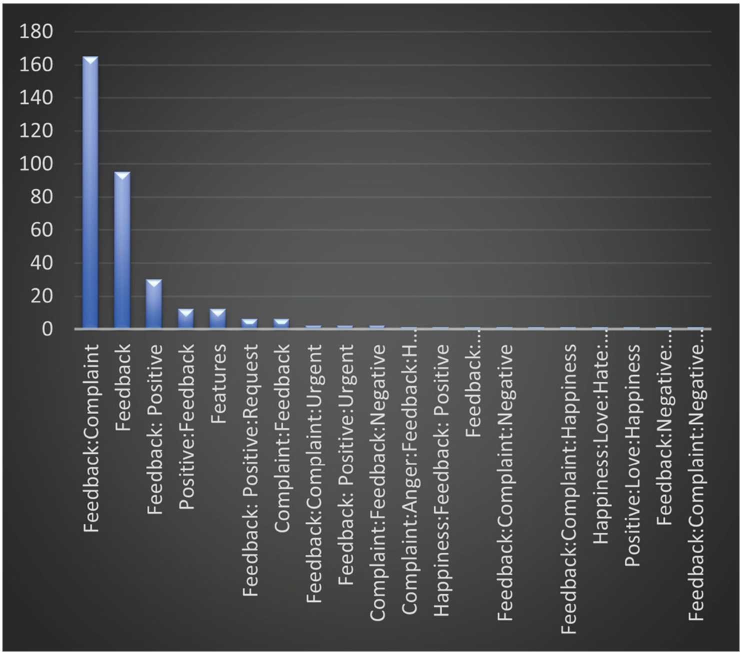

Fig. 1.4 Top 20 categories containing the largest group of tweets

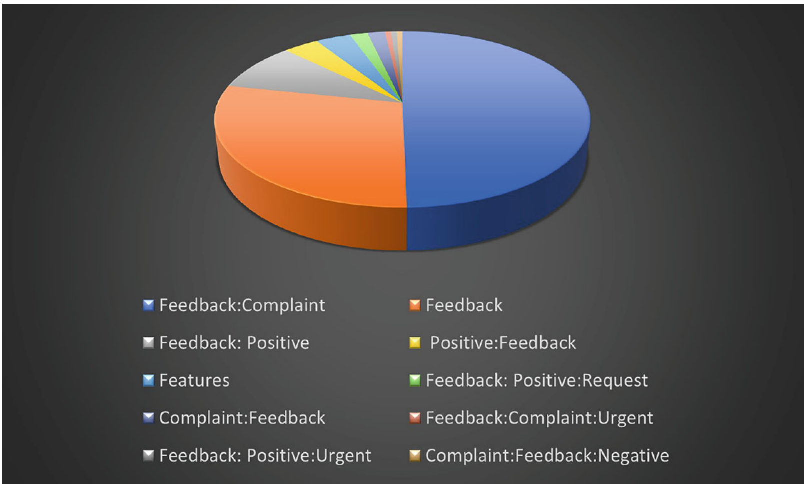

Fig. 1.5 Top 10 categories percentages out of the overall

1.7 Conclusions

By searching for right words and combination of words through the PubMed database, and using a previously trained ISUED model for finding the right words through a dataset of tweets, a semantic analysis can be obtained describing the relation between the required words as described in the previous sections. Medical language used in clinical examinations, books, and literature is highly specialised which requires careful observation and extraction through various text mining techniques, in some cases employing multiple techniques. This is compounded by the fact that patient medical records, test results, prescription, and clinician notes are all free text based in the form of unstructured data. Searching for relevant concepts and identifying their mutual relations is crucial, as it forms the bases of development of AI based applications which can aid in the early detection of disease and better healthcare management of the patient.

References

1. I. of Medicine (US) Committee on Quality of Health Care in America, *"Crossing the Quality Chasm: A New Health System for the 21st Century"*. Washington, DC: National Academies Press (US), 2001.
2. WHO. (2010) "telemedicine: opportunities and developments in member states: report on the second global survey on ehealth". [Online]. Available: https://www.who.int/goe/publications/goe_telemedicine_2010.pdf
3. R. Roine, A. Ohinmaa, and D. Hailey, "Assessing telemedicine: a systematic review of the literature," *CMAJ*, vol. 165, no. 6, pp. 765–771, 2001. [Online]. Available: https://www.cmaj.ca/content/165/6/765
4. M. Systems. (2022) "telemedicine: The complete guide for 2022 | modality". [Online]. Available: https://www.modalitysystems.com/hub/blog/telemedicine
5. T. Care. (2019) "tyto care and bradford partner in nhs 'first' to bring virtual care to children in uk's youngest city". [Online]. Available: https://www.prnewswire.com/news-releases/tyto-care-and-bradford-partner-in-nhs-first-to-bring-virtual-care-to-children-in-uks-youngest-city-300976796.html
6. S. Vijayarani, J. Ilamathi, and S. Nithya, "Preprocessing techniques for text mining-an overview," *Int J Comput Sci Commun Networks*, vol. 5, no. 1, pp. 7–16, 2015. [Online]. Available: https://www.researchgate.net/publication/339529230
7. R. C. of Psychiatrists. (2022) "covid-19: Remote consultations | royal college of psychiatrists". [Online]. Available: https://www.rcpsych.ac.uk/mental-health/treatments-and-wellbeing/remote-consultations-and-covid-19
8. GMC. (2022) "remote consultations – gmc". [Online]. Available: https://www.gmc-uk.org/ethical-guidance/ethical-hub/remote-consultations
9. M. Esmail and M. A. Alzain, "Mobile based tele-medicine diabetic retinopathy screening," in *2018 International Conference on Computer, Control, Electrical, and Electronics Engineering (ICCCEEE)*, 2018, pp. 1–6.
10. K. Donelan, E. A. Barreto, S. Sossong, C. Michael, J. J. Estrada, C. A. B., J. Wozniak, and L. H. Schwamm, "Patient and clinician experiences with telehealth for patient follow-up care," *The American Journal of Managed Care*, vol. 25, no. 1, pp. 40–44, 2019. [Online]. Available: https://www.ajmc.com/view/patient-and-clinician-experiences-with-telehealth-for-patient-followup-care

11. S. Morony, K. Weir, G. Duncan, J. Biggs, D. Nutbeam, and K. J. McCaffery., "Enhancing communication skills for telehealth: development and implementation of a teach-back intervention for a national maternal and child health helpline in australia," *BMC health services research*, vol. 18, no. 1, p. 162, 2018.
12. C.-M. D, L.-V. A, L. C. C, B.-M. R, R.-M. E, R. P. P, and L.-L. R, "A non-randomized clinical trial to examine patients' experiences and communication during telemonitoring of pacemakers after five years follow-up," *PLoS ONE*, vol. 16, no. 12, 2021.
13. Z. Bilgrami, A. Abutaleb, K. Chudy-Onwugaje, P. Langenberg, M. Regueiro, D. A. Schwartz, J. K. Tracy, L. Ghazi, S. A. Patil, S. M. Quezada, K. M. Russman, C. C. Quinn, G. Jambaulikar, D. B. Beaulieu, S. Horst, and J. Cross, R. K., "Effect of telemedicine for inflammatory bowel disease on patient activation and self-efficacy," *Digestive diseases and sciences*, vol. 65, no. 1, pp. 96–103, 2020.
14. D. M. Mann, J. Chen, R. Chunara, P. A. Testa, and O. Nov, "Covid-19 transforms health care through telemedicine: Evidence from the field," *Journal of the American Medical Informatics Association*, vol. 27, no. 7, pp. 1132–1135, 2020.
15. D. CM, S.-C. JM, M. BD, O. KA, G. NE, D. CM, P.-H. L, W. LS, and V. JA, "Determining factors that influence parents' perceptions of telehealth provided in a pediatric gastroenterological practice: A quality improvement project," *Journal of pediatric nursing*, vol. 62, pp. 36–42, 2022.
16. A. Parikh, A. A. Kumar, and E. Jahangir, "Cardio-oncology care in the time of covid-19 and the role of telehealth," *JACC: CardioOncology*, vol. 2, no. 2, pp. 356–358, 2020.
17. N. P. Pilosof, M. Barrett, E. Oborn, G. Barkai, I. M. Pessach, and E. Zimlichman, "Inpatient telemedicine and new models of care during covid-19: Hospital design strategies to enhance patient and staff safety," *International Journal of Environmental Research and Public Health*, vol. 18, no. 16, 2021.
18. R. Harding, J. M. Carrasco, J. Serrano-Pons, J. Lemaire, E. Namisango, E. Luyirika, T. Immanuel, A. K. Paleri, L. Mathews, D. Chifamba, L. Mupaza, C. L. Martínez, L. Zirimenya, M.-C. Bouësseau, and E. L. Krakauer, "Design and evaluation of a novel mobile phone application to improve palliative home-care in resource-limited settings," *Journal of Pain and Symptom Management*, vol. 62, no. 1, pp. 1–9, 2021.
19. C. S. Kruse, K. Lee, J. B. Watson, L. G. Lobo, A. G. Stoppelmoor, and S. E. Oyibo, "Measures of effectiveness, efficiency, and quality of telemedicine in the management of alcohol abuse, addiction, and rehabilitation: Systematic review," *Journal of medical Internet research*, vol. 22, no. 1, p. e13252, 2020.
20. M. A. Shareef, V. Kumar, Y. K. Dwivedi, U. Kumar, M. S. Akram, and R. Raman, "A new health care system enabled by machine intelligence: Elderly people's trust or losing self control," *Technological Forecasting and Social Change*, vol. 162, p. 120334, 2021.
21. T. T. Sudmann, E. H. Fredriksen, I. T. Børsheim, and I. Heldal, "Knowledge management from senior users of online health information point of view," *The Electronic Journal of Knowledge Management*, vol. 18, no. 3, p. 325–337, 2021.
22. C. S. Kruse, P. Karem, K. Shifflett, L. Vegi, K. Ravi, and M. Brooks, "Evaluating barriers to adopting telemedicine worldwide: A systematic review," *Journal of Telemedicine and Telecare*, vol. 24, no. 1, pp. 4–12, 2018.
23. S. Bakshi, U. Tandon, and A. Mittal., "Drivers and barriers of telemedicine in india: Seeking a new paradigm," *Journal of Computational and Theoretical Nanoscience*, vol. 16, no. 10, pp. 4367–4373, 2019.
24. M. Trovati, J. Hayes, F. Palmieri, and N. Bessis, "Automated extraction of fragments of bayesian networks from textual sources," *Appl. Soft Comput.*, vol. 60, no. C, p. 508–519, nov 2017.
25. M. Trovati, H. Zhang, J. Ray, and X. Xu, "An entropy-based approach to real-time information extraction for industry 4.0," *IEEE Transactions on Industrial Informatics*, vol. 16, pp. 6033–6041, 2020.