PUBLISHED VERSION

Haneen Banjar, David Adelson, Fred Brown, and Tamara Leclercq

Personalized medicine support system : resolving conflict in allocation to risk groups and predicting patient molecular response to targeted therapy

---

---

http://hdl.handle.net/2440/111210

PERSONALIZED MEDICINE SUPPORT SYSTEM: RESOLVING CONFLICT IN ALLOCATION TO RISK GROUPS AND PREDICTING PATIENT MOLECULAR RESPONSE TO TARGETED THERAPY

Haneen Banjar[1, 2], David Adelson[3,] Fred Brown[2] and Tamara Leclercq[4]

[1]School of Computer Science, University of Adelaide, Adelaide, Australia.
[2]The Department of Computer Science, King Abdul Aziz University, Jeddah, Saudi Arabia.
[3]School of Molecular and Biomedical Science, University of Adelaide, Adelaide, Australia
[4]Cancer Theme, South Australian Health and Medical Research Institute (SAHMRI), Adelaide. South Australia, Australia

ABSTRACT

Treatment management in cancer patients is largely based on the use of a standardized set of predictive and prognostic factors. The former are used to evaluate specific clinical interventions, and they can be useful for selecting treatments because they directly predict the response to a treatment. The latter are used to evaluate a patient's overall outcomes, and can be used to identify the risks or recurrence of a disease. Current intelligent systems can be a solution for transferring advancements in molecular biology into practice, especially for predicting the molecular response to molecular targeted therapy and the prognosis of risk groups in cancer medicine. This framework primarily focuses on the importance of integrating domain knowledge in predictive and prognostic models for personalized treatment. Our personalized medicine support system provides the needed support in complex decisions and can be incorporated into a treatment guide for selecting molecular targeted therapies.

KEYWORDS

Personalized Medicine Support System, Molecular Targeted Therapy, Predict Molecular Response, Risk Assessment & Cancer Treatment.

1. INTRODUCTION

1.1 Personalized medicine

Personalized medicine could be used to establish molecular characteristics that are unique to an individual, usually because of varying genetic makeup. With this, doctors can offer a more specific diagnosis andeffective treatment protocol for an individual's disease. In clinical decision-making, personalized medicine is used to make decisions that maximize the outcomes and minimize the side-effects of treatment by using available knowledge about an individual [1]. Many studies have focused on personalized care, and its scope of application has increased because it leads to more successful outcomes and minimizes the chances of adverse reactions associated with certain treatment plans. Personalized medicine involves classifying individual

patients into sub-populations that pose unique reactions, susceptibilities or general responses to particular treatments to determine the best treatment approaches for every patient [2].

Personalized medicine offers high precision, and it has attracted much interest because it affords advantages such as improved healthcare and reduced need for the development of new medicines [3]. Personalized medicine also helps to minimize unnecessary costs by reducing both the time spent on treatment and the failure rates during clinical trials [4].Economic value increases when caregivers implement personalized medicine in healthcare, because it helps to reduce the unnecessary use of resources.

Traditional personalized medicine contrasts with modern personalized medicine in how suitable therapies are identified for patients. In the former, a diagnosis is predicted based on a patient's family history, social circumstances, environment and lifestyle. In the latter,this prediction is based on a patient's genetic makeup [2]. Medical researchers continue to emphasize that their studies are updated with the most effective treatment protocols used to treat cancer patients. However, different patients show different responses to the same therapy. For example, when subjected to the same therapeutic protocol, the clinical outcomes of two patients may be different in terms of response and survival[5]. This undesirable situation could potentially be avoided by improving predictive assays to predict an individual's response to therapy at diagnosis. Indeed, during recent years, studies have tried to individualize medicine by using predictive factors to determine initial treatments. Many studies have used different techniques to investigate factors that could affect drug responses. Sioud et al. [6]noted that individualized treatment algorithms could depend on integrating the factors that influence drug responses. For example, studying a single biomarker as a predictor could indicate the response before treatment and predict the risk to the individual [7]. Another approach involves predicting a drug reaction in terms of toxicity or resistance by using an individual's genotype data and clinical data to improve the individual's care [8].Inpharmacogenetics, an individual's response to a specific therapy is studied based on his/her genotype information. The European Science Foundation established a roadmap for developing personalized medicine [9] and stated that to build a predictive model, data for individualized treatment options were needed, and response variations in different sub-populations of patients and diseases needed to be stored. Fagan and Shortliffe[10] noted the power of using a computer to (1) analyze the combination of biological and clinical data, which are disparate data sources, and (2) discover the relationships amongst complex database schema that store medical records.

As treatment optimization is an important medical advancement, many computer science researchers have focused on predicting the response to treatment [11]. Indeed, many studies have reported on the development of a prognostic or predictive model that could predict the side-effects of drugs [12-15]; such models use intelligent techniques to address medical issues. This gives researchers a new way of applying the strength of machine-learning techniques to real medical problems. Current intelligent systems could potentially aid in transferring advancements in molecular biology into practice, especially for predicting molecular response to targeted therapy and the prognosis of patients (in different risk groups) in cancer medicine.

1.2 Clinical Decision Support System

A clinical decision support system (CDSS) is a type of decision support system that analyses medical data to assist healthcare providers in making appropriate and automated clinical decisions [16]. Constructing a CDSS requires substantial modeling activity by selecting relevant medical data and a problem-solving strategy to reach appropriate conclusions. For the analysis, collaboration with clinical experts is needed to model relevant application areas. Based on the nature of this collaboration, there are two types of CDSSs: knowledge-based CDSS and non-

knowledge-based CDSS [17]. A key research area in CDSS is the advancement and application of knowledge-based systems because of the usability of expressing various types of human knowledge in an intelligent system. Combining expert knowledge and discovered knowledge in the medical domain maximizes the qualities that they have separately [18]. There are also three classificationsfor CDSS: information management, situational awareness and patient-specific[19]. Information management uses buttonsfor obtaining up to date information; situational awareness uses alerts or dashboard for obtaining information; and patient-specific may be implemented for different purposes such as diagnosis, treatment management or recommendation [20]. Identifying the key needs and functional requirements and determining how to evaluate the system are interrelated issues in designing and implementinga CDSS [21]. Understanding a medical problem and its domain are also primary requirements for implementinga knowledge-based CDSS. Additional requirementsinclude knowing who will deliver the information from the CDSS and how the CDSS can provide support starting from diagnosis and selecting treatment to monitor and follow-up the outcomes.

In the last decade, information technology has been introduced to improve clinical practice [22]. Information systems make data storage and data retrieval easier. This enhances the use of CDSS in clinical practice, and it helps to prevent errors and deliver needed information at the time of request. The development of CDSS has been influenced by the purpose of use and advancements in medical technology. Researches are motivated by the concept of personalized medicine, and they have introduced personalized modeling for developing CDSSs based on the information of individual patients, such as clinical and molecular information [23].

Automatic knowledge acquisition techniques have also attracted much research interest. Knowledge acquisition is the first step in building a knowledge base, and it requires techniques to drive knowledge. Data mining is widely used for knowledge acquisition,and the knowledge discovery process is used as a schema to automatically discover knowledge, which can include the shape of patterns or associations among data features and rules. This process generally includes four steps: problem understanding and data understanding, data pre-processing, model induction, and post-processing [24]. A recent review [25]discussed knowledge discovery in medicine. In this review, the authors identified the primary studies in medical research applications through data mining techniques. They found that the focus has moved to the medical domain more than in the past and that the use of hybrid systems is the future for solving complex problems in medicine. Complex problems in the medical domain began with the success of sequencing the human genome. The morenew biological information was made available, the greater the demand to build intelligent systems that can utilizethis information. Intelligent systems such as CDSSs should be adaptive to any new information made available on an ongoing basis. From adaptive behavior, evolving systems have been introduced to use particularly intelligent systems and online learning algorithms for knowledge acquisition from data and for realizing advanced model structures in data mining tasks and parameters [26].

1.3 Knowledge-based Systems

Researchers have also focused on the integration of knowledge from multiple sources, which is a key issue in many areas such as collaborative knowledge systems, group decisions and distributed expert systems, where knowledge from multiple sources is often contradictory. We need to be comfortable with a wide range of knowledge resources that will help clinicians in decision-making, as a conflict between two sources, rules, or decision-makers is often observed in real cases.

Conflicts in knowledge-based systems can be divided into three types: schema conflicts, data conflicts, and knowledge conflicts [27]. Schema conflicts can arise from the use of different schema definitions such as tables or objects. Data conflicts may ariseowing to incorrect data. A

study compared the ability of several techniques to reduce inconsistency in data [28]. Knowledge conflict may arise when multiple sources of knowledge are integrated. Knowledge conflict is a major issue in distributed artificial intelligence systems [29]. Adrian et al.[30] defined contradiction as an 'inability for all conceived statements or beliefs to be simultaneously true'. A contradiction may also be referred to as an inconsistency. For instance, there may be a contradiction between sentences such that one sentence must be true and the other must be false.

This common situation requires some advanced methodologies to resolve the conflict. One active research direction is the use of the multiple criteria decision analysis (MCDA) approach [31, 32]. This technique is developed to capture the relative preference information of decision-makers who are involved in a conflict. There is a prescriptive approach based on progressive preferential knowledge [33]. This approach resolves the conflicts of rules in a knowledge-based system by using decision analysis techniques, where incorporating a user's preference in judgment about the rules is important. These methods for conflict resolution in a rule-based system always need toincorporate a user's preferences in some specific problem domain that is very difficult to obtain in the medical domain. Experts' opinions, including preferences, are often vague and difficult to estimate using exact numerical values. Noor-E-Alam et al. [34] developed a framework that used qualitative form as a linguistic term to represent experts' opinions. Their approach aimed to use multi-expert multicriteria decision making (ME-MCDM) to reach a single decision by integrating multiple experts' opinions. The linguistic truth value (LTV) can be used to judge alternatives for ME-MCDM. Two suitable algorithms are used to handle conflict aggregation: possibility measure and averaging conflict aggregation. Junming et al. [35] designed a framework using fuzzy case base reasoning for conflict resolution. This approach aimed to find similar cases in the resources and to retrieve the information that resolved the problem. They defined linguistic variables for comparing the retrieved information. Sometimes, it is important to measure the inconsistency; based on the results, the researcher can decide what to do with it. For example, one can measure the conflict and agreement between two knowledge bases [36] or the inconsistency of knowledge bases [37-41].Some studies used machine learning techniques to resolve conflicts. Recent studies on multi-agent systems aimed to resolve conflicts by combining machine learning techniques such as Bayesian network, case-based reasoning and expert systems [42]. This study demonstrated that 'the choice of the specific technique for a given domain depends on the specification of the domain'.

1.4 Medical Data and Knowledge

Clinical trials are usually conducted in a series of phases. Each clinical trial phase aims to answer a specific research question. Phase I is designed to test a new treatment in a small population and to identify side-effects and treatment safety. In Phase II, researchers test the drug on a larger population (several hundred individuals) and evaluate its efficacy and safety. In phase III, the trial is designed to study the efficacy in a population of several hundred to a thousand individuals. Phase IV is designed to monitor the drug in the general population after it is marketed[43]. Late phases of the clinical trials,such as phases II, III and IV, are considered important sources of information that can be used to build mathematical models. There is a rapid increase in the number of electronic medical research databases that provide an opportunity for researchers to reuse medical data to create mathematical models. To access the clinical trial data, the NCI[44]is a US agency that lists ongoing clinical trials of molecular targeted therapies. Investigators within hospitals and medical centers conduct most of the NCI-supported trials. The NCI provides the full trial description and the name of the principal investigator for these studies. Researchers can contact the investigators and collaborate with them.

Clinical trial data may be biasedin several aspects: sampling, referral, selection, method and clinical spectrum. The clinical trial may use sampling methods, sample size and inclusion and exclusion criteria. Another aspect is the referral bias, where a specialist refers patientsand thus the

data represents pre-selected patients who have high prevalence of a disease. The selection bias is clear when the clinical trial data includes a group based on various demographics. Data may be collected using different measurements, leading to different precisions and specifications. Finally, the clinical spectrum bias is represented in the patient's record, which may show other medical problems with the disease [24]. For instance, Saussele and Pfirrmann[45] reported clinical trials of chronic myeloid leukemia (CML). They noted several issues that may challenge the reuse of a clinical trial data. According to Saussele and Pfirrmann, the definition of 'remission'varies in clinical trials depending on the major molecular response (MMR) or complete cytogenetic response (CCR). In addition, clinical trials use different primary endpoints such as 12-month MMR or 12-month CCR to determine the treatment success.

From patient care to patient administration, electronic health records (EHR) are reused in many studies to answer specific research questions [46-49]. Cases were matched with enquires based on obtained research criteria for patient inclusion, and a dataset of many matches can be generated for analysis. The EHR may include sparse data or missing values, as some patients may not seek frequent care. The EHR quality would likely impact the bias in the research finding or the modeling performance. Thederivation of key variables is also an important aspect when dealing with EHR, as the values may be recorded in different ways in different systems. This arises because of varying definitions between sources. The data quality and correct values of derived key variables will concern researchers,and many algorithms can be investigated during pre-processing to improve the data quality,thereby providing reliable results [50].

The reuse of medical data and knowledge for cancer treatment requires advanced computational methods such as intelligent systems. Using multiple sources of data and knowledge for personalized medicine support systemswillcreatelarge amounts of information to deal with for evaluating therapies and potential diagnostic and prognostic markers.

1.5 Predictive vs. Prognostic

In oncology, predictive markers differ from prognostic markers [51, 52]. Predictive markers or factors used to evaluate specific clinical interventions, and they can be useful for selecting a treatment because they can directly predict the response to a treatment.Prognostic markers are used to evaluate a patient's overall outcomes, and can be used to identify the risk or recurrence of a disease. By using advanced molecular technologies, many predictive and prognostic factors have been introduced in cancer medicine. Studying predictive molecular biology in detail is beyond the scope of this study;however, the list of possible predictive factors that have been published previously in separate studies such as clinical, pathological and molecular tests are the primary focus of our study. Based on the available data from medical research and clinical trials, one can create a list of possible predictive factors and prepare it for further investigation and also explore any relationship among the predictive factors.

Several predictive and prognostic tests based on molecular biology have revolutionized cancer medicine. In our study, we propose collecting information about possible predictive factors and prognostic scores that are published in top-ranked medical journals such as Science, Medline and PubMed journals and in other publications indexed under keywords such as 'prediction of molecular response', 'predict outcomes', 'prognostic model or score', 'prognostic indicators' and 'predictive factors'. We propose communicating with several clinical trial investigators who are studying the efficiency of molecular targeted therapies and patient responses by using several predictive and prognostic factors for decision-making. Predictive factors that are regulated for use should be used. According to Mehta[51], predictive markers that are commercialized for decision-making should be reviewed by the FDA and be certified with adequate evidence. Official and ethical approvals on the reuse of medical research data should be obtained before establishing a list of predictive and prognostic factors from clinical trials.

1.6 Molecular Targeted Therapy

Targeted cancer treatments are also called 'molecularly targeted drugs', 'molecularly targeted therapy' and so on [53]. Molecular targeted therapy is a sort of customized therapeutic treatment that aims to treat cancer by acting on unique mutations, overactive kinases, or protein anomalies that drive cancer development. Specifically, the medications used for targeted treatment are intended to act on a specific biochemical pathway vital to the survival of that specific cancer [54]. The FDA has approved numerous targeted cancer treatments to treat specific types of cancer [55]. Others are being examined in clinical trials (human testing), and numerous more are in preclinical testing (animal testing) [56].The National Cancer Institute (NCI) website lists multiple targeted therapies that have been approved for particular types of cancer[57]. For example, two targeted therapies named Bevacizumab (Avastin®)and Everolimus (Afinitor®)have been approved for treating brain cancer. Multiple targeted therapies are also available for leukaemia patients. To personalize targeted therapies, we need to determine the best therapeutic agent for an individual patient at the time of diagnosis. This can be done by studying the efficacy of each targeted therapy and developing a clinical trial. The NCI website provides a list of clinical trials of targeted therapies[57]. These trials are an important resource for shifting information from the laboratory to the implementation of personalized medicine. Developments in molecular technology will play a major role in realizing predictive, preventative and personalized medicine. Here, we also focus on the issue of how to reuse patients' existing profiles, including personal molecular information gathered in clinical trials for resolving conflict in allocation to risk groups and predicting patient molecular response to targeted therapy.

1.7 Research Motivations

To benefit from molecular technology data for treatment and monitoring progress in cancer patients, implementing an intelligent system that integrates medical knowledge, identifies relevant features and suggests drugs or treatments for patientsas well as adapting a new source of data and knowledge to change, remains a future challenge. In this research, we focus on the following main question: 'How can one develop a personalized medicine support system that allocates risk groups and predicts the molecular response tomolecular targeted therapy, where consistency in decisions is paramount?'

Therefore, previous studies have not considered improving knowledge-based CDSS that has a high-quality knowledge base to provide consistency in decisionsmade based on a patient's molecular information for predicting his/her molecular response.The main objective of this framework is to build a personalized medicine support system as a knowledge-based system changes and evolves over time. This intelligent system will provide clinicians, patients and researchers with a platform to extract knowledge and to improve the personalized treatments of patients. Framework functions have not been considered in previous studies on personalized medicine,and they have also not been specifically applied in medical literature to the problem we are considering.Doing so will add substantial value to the work. Ageneric framework is presented that includes the following main sub-functions: (1) consistency test between scoring systems for predicting outcomes in cancer patients treated with molecular targeted therapy, (2) resolving conflicts in validated prognostic scores, (3) developing a predictive model to predict molecular response to molecular targeted therapy, (4) select relevant predictive factors for predicting molecular response to molecular targeted therapy, (5) derive a clinical prediction rule that uses clinical, molecular and cell count observations (these predictive factors were collected atdiagnosis and categorized based on domain knowledge) and (6) adaptive methods that can evolve knowledge-based systems over time.

2. PROPOSED FRAMEWORK

In this section, the system overview ispresented, stages of implementing a personalized medicine support system are discussed and the system structure is described in detail.

2.1 Overview of Personalized Medicine Support System

Patient treatment has become increasingly difficult in recent years as various molecular targeted treatment options have become available. Clinicians aim to treat individual cancer patients with the most appropriate therapy. The ultimate challenge is to decide the most appropriate treatment strategy for an individual patient. To facilitate this, the disease assessment of an individual patient in terms of risk profile and molecular response to initial molecular targeted treatment should be determined. In prognostic models, low-risk patientsare preferentially treated with the least toxic and safest therapeutic options. These patients' molecular responsesare predicted using the predictive model built for the initial molecular targeted treatment. In higher-risk patients, higher-toxicity or combination treatment options may be favoured. High-risk patients or those who fail to respond to initial therapy will benefit from this approach as their response should bepredicted, thereby preventing resistance and/or intolerance to therapy, that may have otherwise ensued. Thus, this system aims to enhance individual disease and treatment management.

A personalized support system is designed for selecting patient therapy. Reused medical research data (clinical trial data) and prior knowledge collected from literature are used with advanced computational techniques to allocate risk groups and predict patients' molecular response to molecular targeted therapy at diagnosis. The personalized medicine support system uses two models as a guide for treatment selection (Figure 1):(1) prognostic model and (2) predictive model.
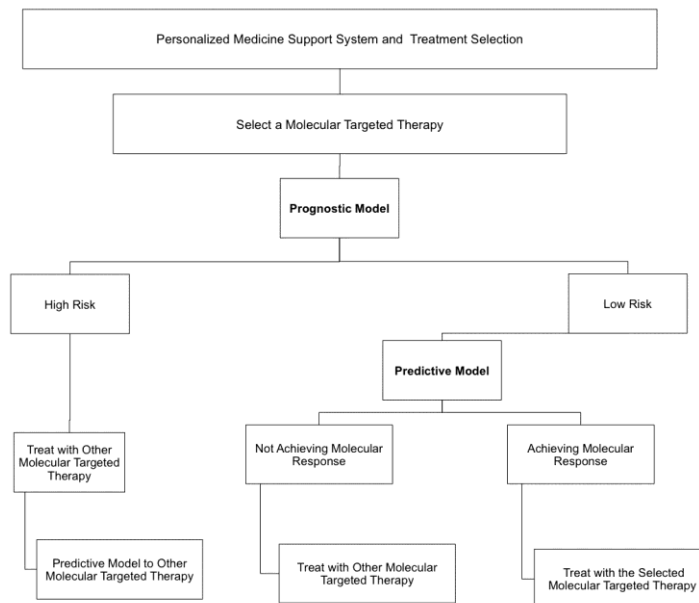

Figure1. Process flow of personalized medicine support system

2.1.1 Prognostic Model

The prognostic model uses four data mining processes: problem understanding and data understanding, data pre-processing, model development and post-processing. In medicine, it is

important to investigate current validated prognostic factors and score systems that are used to stratify cancer patients according to risk profile to ensure appropriate treatment. Historically, prognostication has seen rapid developments, and various scoring systems have been developed to optimize the use of clinical experience in cancer treatment. Our system tests the consistency between these scores while identifying risk categories, and the main procedure in this step is identifying whether conflict groups exist. During data pre-processing, prior knowledge is prepared for the analysis, and necessary prognostic factors are determined to implement previous prognostic scores. The data store contradiction in prior knowledge (validated prognostic scores) is prepared in this process. Feature ranking, conflict analysis and data normalization on a unified scale are performed to help in describing the relation between prognostic scores.

Model development combines validated scoring systems to determine whether the combined model may further improve the prediction compared with a single source of knowledge. In this process, one needs to know what patients' outcomes are as conveyed by previous prognostic scores. Using combined methods to resolve conflict in conflict groups adds another dimension of knowledge. Three levels are used for combining the validated prognostic scores and factors: combination, classification and feature. Prognostic factors are combined in the feature level,and prognostic scores are combined as score values and risk categories. Evaluating the combined methods and selecting the best approach can eliminate contradictions. Highest-accuracy methods will be selected for external evaluation. The final process is to use the selected prognostic models and validation on unseen data. Extending the prognostic model to adopt a new prognostic score or prognostic factor requires an additional process. Figure 2 shows the process flow.
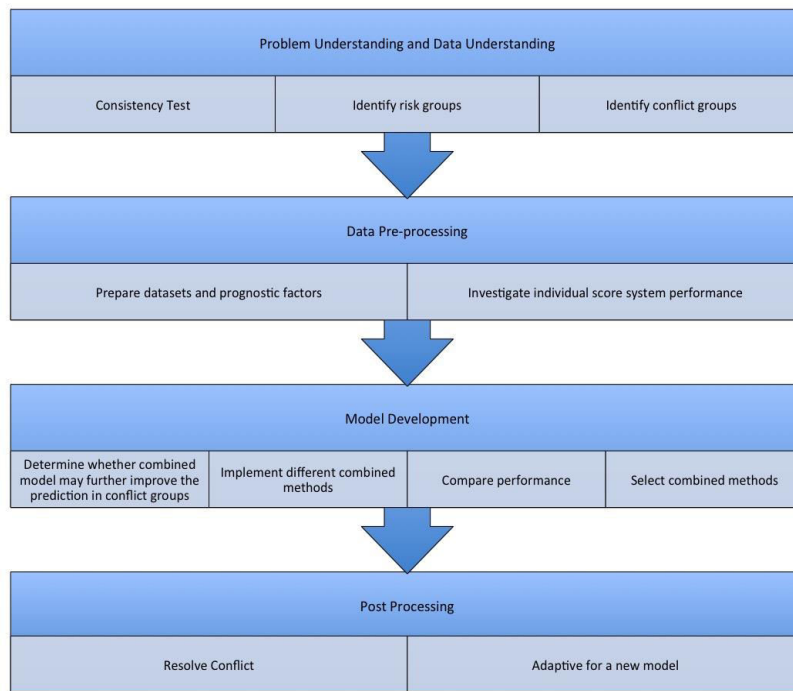

Figure2.Process flow of data mining processes used in prognostic model

2.1.2 Predictive Model

The predictive model also uses four data mining processes: problem understanding and data understanding, data pre-processing, model induction and post-processing. Figure 3 shows the process flow. The first process in knowledge discovery is divided into two stages. In the first

stage, each predictive factor used as an input parameter, studied to understand whether it is relevant to the prediction. When medical research data is reused, it is important to understand the data, such as the comprehensive treatment protocol or specifications for responses.Inclusion and exclusion criteria should also be decided before conducting any analysis to ensure that the material meets clinical requirements for the induction of a model with an accurate population.The second phase is data pre-processing. A statistical method is used to input missing values in predictive factors and prognostic scores from prior knowledge. Removing these patients is not an option in medical research data because most medical research data comprises small populations. For instance, patients who are enrolled in clinical trials mustprovide signed consent to undergo molecular targeted therapy in phase II clinical trials, as the safety profile of the drug is still not confirmed. Furthermore, we use domain knowledge to reformat the predictive factors. This process is performed before model induction by using medical knowledge to split continuous predictive factors into two or more linguistic terms and intervals. Thus, the clinical prediction rules can be interpretedmore easily based on domain concepts.
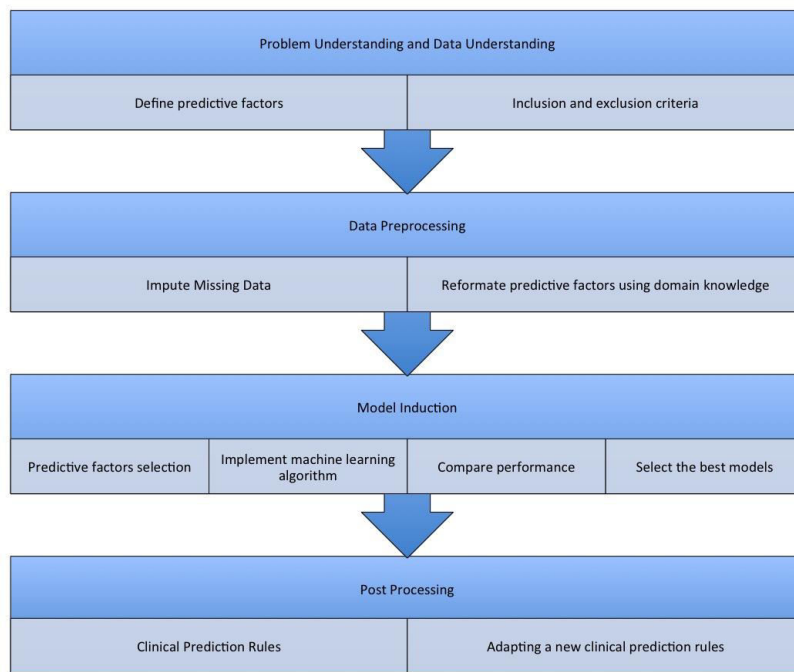

Figure3.Process flow of data mining processes used in predictive model

The third phase is model induction. First, there are three general steps for predictive model induction: training set, learning algorithm and performance evaluation on testing set. The training set represents the patient profiles that store the collected molecular, clinical and blood count information for individuals. We use available factors irrespective of whether we have prior knowledge that a factor's ability to predict patient response has been identified or whether a factor is new in the field. This is because these factors will be used in feature selection techniques [58] to reduce the dimension of the data. Feature selection (predictive factor selection) is important to minimize the cost of the requested tests for a patient. In addition, technically, it can increase the model accuracy because removing unnecessary data may increase the quality of the training set [59]. The second step is the learning algorithm.This is an important step in which the selected machine learning techniques are used for learning [60]. The selected algorithm should outperform previous methods and should be interpretable. Interpretability is an important aspect for medical experts. The selected machine learning algorithm is popular in the medical domain as a primary analytical model for discovering the relation among variables, dividing a dataset into

groups based on shared characteristics (classification) and providing a set of clinical prediction rules that can be applied to new patients. The model performance is evaluated by comparing patients' actual outcomes (molecular response) with those predicted by the model. A confusion matrix can help in measuring the performance based on accuracy, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), G-mean(geometric mean) and F-score(weighted harmonic mean of sensitivity and PPV). Two sets are used in the model induction processfor validating a small dataset: training and testing setand cross-validation set[61]. When a dataset is divided into a number of equal folds, one fold is used for validation and the others, for training. The process is repeated to ensure that all folds are used one time for validation. The standard deviation is measured after completing this cross-validation approach.

In post-processing, we compare the performance of the predictive models using validation sets and select the model that achieves the highest validation performance. The predictive model generates clinical prediction rules and discovers unknown knowledge by selecting relevant predictive factors (molecular tests, clinicalbase and cell count observations collected at diagnosis and categorized based on available knowledge) and extracting the relationship between the predictive factors and the molecular outcomes.This would prove that the approach could be applied in practice for treatment management by integrating domain knowledge in the learning process.A machine learning algorithm is used to discover unknown knowledge and perform classification. The predictive models present interesting properties of the data and use historical data to predict the behavior of unseen data.

2.2 Prognostic and Predictive Factors

The factors can be selected based on experts' choices and publications. All predictive factors and prognostic factors have an equal chance of being studied by the selected feature selection techniques to compare the performance of each or group of predictive factors to predict the molecular response, and prognostic factors to allocate risk group. Many factors may influence the correlation between baseline data and end-point molecular response.There are manycategories of factors that maybe clinically based, biologically (molecularly) based, environmentally based, assay results and family history data etc. In a personalized medicine support system, we collect information from literature and create possible predictive and prognostic factors based on available data. We select the predictive and prognostic factors that include the patient's individual information. For each patient$p_i has pf_z = \{pf_1, pf2, ..., pfz\}$, where$pf_z$ is the predictive and prognostic factors, z is the total number of factors, $i = \{1, 2, 3, ..., n\}$is the number of patients and n is the total number of patients in dataset D. Finally, D contains rows of patient $P_i$and the columns arepf_z. Ideally, the predictive and prognostic values should have been previously studied and validated by using a group of patients with known responses to molecular targeted therapy.The following are possible factors that can be used as inputs for the system:

- Clinical Factors: clinical factors are any specifications or measurements related to the patient, such as spleen size or agethat can be observed or obtained by clinicians. Clinical factors are collected on the day of diagnosis.

- Biological Factors: patients are not biologically identical. Biological factors include blood cell counts or the activity levels of molecular parts. Multiple biological factors may influence the treatment response, and these may not be directly related to treatment outcomes. In practice, pathological tests are considered a primary pre-treatment step for individual patients. Including an individual's information helps in realizing more specific treatment. In addition, currently, most interesting developments are related to real-time quantitative polymerase chain reaction (RQ-PCR) analysis for DNA sequences. To

monitor the molecular response, RQ-PCR, a sensitive technique, is used to quantify the level of mRNA transcripts in the peripheral blood of patients.

- Predictive Assays: assays are used to predict possible outcomes following treatment [62]. Predictive assays can guide a clinician in making a rational choice of therapy at diagnosis. Identifying predictive assays, especially genetics-based ones, in a diverse group of individuals may provide substantial knowledge about the mechanisms of individual differences in responses to molecular targeted therapy.Groups whose assays show similar results can be treated similarly. Studying significant differences by the statistics of different thresholds for identifying groups of response and using a threshold less than a cut-off point (e.g. $p < 0.5$) are commonly used approaches in medical study. Here, we used a predefined threshold from publications and applied it to our dataset.

2.3 Inclusion and Exclusion Criteria

Inclusion criteria identify the base standard for including patient data in the study,and exclusion criteria exclude patient data based on predefined requirements. The reuse of medical research data requires refinement of patients in the secondary analysis. Inclusion and exclusion criteria help researchers in optimizing existing medical data to make it suitable for new research. In clinical trial data, patients adhere to research protocols and some standards. This clinical trial data usually uses information available for reuse in a second analysis if it first satisfies a new study's inclusion and exclusion criteria. Understanding data is important to obtain accurate results and establish effective research. Building personalized medicine, as a predictive model to predict molecular responses to molecular targeted therapy should follow these criteria in the reuse of medical research data:

- In clinical trials of molecular targeted therapy, patients from different trials must follow the same treatment protocol.
- Molecular responses must be monitored according tointernational standards based on a pre-identified treatment guide.
- Input data should be filtered at the diagnosis or pre-treatment stage.
- Output prediction end-point should be identified.
- If there is a disease phase, patients should be in the same phase or stage.

Some clinical trials provide useful information by changing the values of some stored data under the guidance of human experts. For example, a patient group is switched to a second-line treatment during the trial. This group may have responded unsuccessfully to the frontline treatment and will therefore be defined as a negative responder to the selected treatment.

The previous list of inclusion and exclusion criteria vary from study to study based on the specific research question under investigation and how much information can be analyzed from the existing resources by collaborating with the domain expert.

2.4 Reformat using Domain Knowledge

Defining the prognostic and predictive factors and reformatting them using domain knowledge is an important process as the final rules' interpretability will be based on the earlier categories. We reformatted the factor values of text, numeral or mixed-type stored data by using existing knowledge such as standard boundaries of blood counts, domain knowledge of clinical expertise, risk categories and previous medical publications. Although some machine learning techniques can handle continuous predictor values, reformatting the data by categorizing each factor in the dataset into subgroups can help improve the comprehensibility of the final model. For each

predictive factor,we reformat the values of factors into categoryC_d,where$d$ is the number of subcategories that represent values in the range using existing knowledge and confirmation by experts in the field.For example, we categorized the value of $pf = IC50$ intofour categories,$d = 4$: $C_1 \leq 0.5$μMasGroup 1, $C_2 > 0.5$ $and \leq 0.7$μMas Group 2, $C_3 > 0.7$ $and \leq 0.95$μMasGroup 3 and $C_4 > 0.95$μMas Group 4. Then, we reformatted the index number of the category. If the final model selects a relevant predictive factor, we can use these categories to distinguish the predictive group on test patients.

2.5 Modeling

The framework is designed to deal with prior knowledge and extracting knowledge from a new source of data. Here, prognostic model used the prior knowledge that is collected from existing prognostic score systems while predictive model used medical data to extract knowledge in the form of clinical prediction rules.

2.5.1 Dealing with Prior Knowledge (Prognostic Model)

The development of the prognostic model was divided into two stages: (1) primary analysis and (2) combined model development.Figure 4. shows the schema for the prognostic model. In primary analysis, available prognostic factors, scores and models are surveyed.
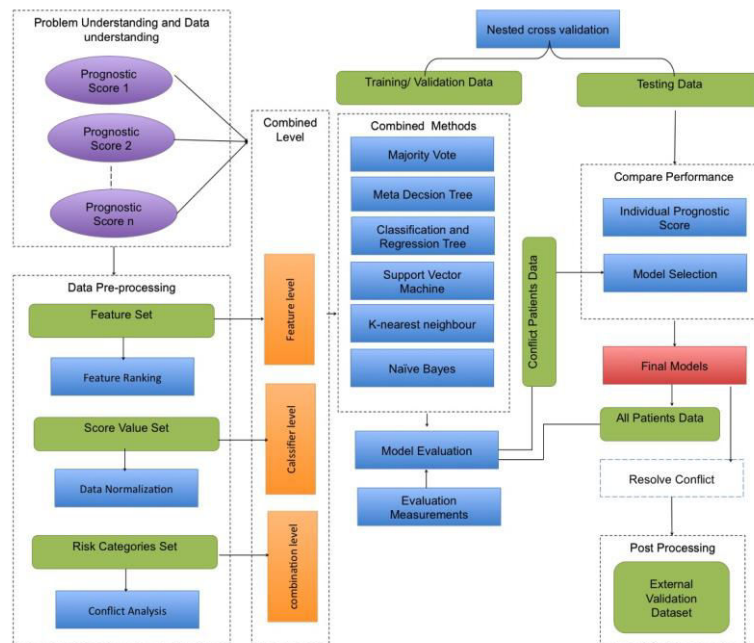
Figure 4.The schema for the prognostic model.

An information matrix is created by adding the validated prognostic method in rows and columns representing the available information from each method. Primary analysis involves investigating the consistency between the prognostic methods by identifying the obtained risk outcomes using this prognostic method. Inconsistency occurs when two different risk categories are applied to the same patient,and it is observed that one prognostic score classifies the patient in one group and the other one contradicts the first classification. The consistency test can be briefly described as follows:

1. Calculate the risk outcomes using prognostic factors included in the prognostic score equation (if available) for all patients in dataset D.

2. Create dataset DPcontaining rows of patient $P_i$and the columns arePS_m, where PS is the prognostic score methodsand m the number of validated prognostic methods.
3. Repeat step (1) for all m.
4. Create an information matrix$IM_1$for risk outcomes from DP to identify the number of risk groups resulting from step (1) calculation.
5. Analyze the risk categories obtained fromIM_1, where R_1 is the minimum and R_2 is the maximum number of categories obtained from all prognostic methods.
6. If R_1 is equal to R_2, go to step (11).
7. CombinedR_2 risk groups to be equal to R_1 groups, where the groups can include all possible combinations of risk groups from R_2.
8. Evaluate the accuracy performance after combining risk groups.
9. Select the combined risk group with the highest accuracy.
10. Combine the risk categories to be equal toR_1in the PS_mthat have more than R_1risk groups.
11. Test consistency in risk outcomes obtained from prognostic methods for each P_i in DP.
12. List in the information matrix $IM_2$the possible combination of risk outcomes, which the number of rows is R_1 raised toPS_m.
13. If $PS_1 = PS_2 = ...PS_m$results have similar risk categories, P_i is considered in a consistent risk group. However, if at least one PS_m has a different risk outcome for the same patient, P_i is considered in a conflict group.

Moving to the second stage of combining models, the steps briefly described as follows:

1. In step (1), three possible data types could be generated:

- DP$_1$: risk categories set as categorical data type resulting from validated risk groups in prognostic model outcomes.
- DP$_2$: score values set as a continuous data type resulting from applying the equation used to calculate the score for the prognostic method.
- DP$_3$: prognostic factors (feature set) as both data types (categorical or continuous).

Normalize DP$_2$ and DP$_3$: Normalization involves mapping the data values into the interval [0-1], where the minimum value is 0 and the maximum value, 1[63]. Normalization is an important process in this study for several reasons. First, the data from prognostic scores is usually not of the same scale. Grouping and comparing patients into logical descriptions using different intervalsin the problem space could make it difficult for the learning algorithm(especially the clustering base algorithm) to learn. Therefore, arranging the data into logical groups based on a unified scale among scores can help in describing the relation between risk groups. Second, for visualization, normalized prognostic score data is easy to represent in the space dimension. Finally, pre-processing real values into scaled values could help in finding a learning function from a space, such as finding the solution using the Euclidean distance between samples.

2. Select the strategies to combinemodels (prognostic methods): model selection or model fusion[64].

- Model selection:Each model is supposed to be an expert in a specific domain of the feature space, and the selected model decides the output of the ensemble.
- Model fusion: Each model is supposed to have complete information on the whole feature space, and we apply combiners to all outputs from the systems.

3. The performance of single PS_m would be the same in model selection. Therefore, Select the classifier fusion method[64], and two possible combiner approach can be used:

- Non-trainable combiner (data-independent) such as majority voting.
- Trainable combiner (data-dependent) methods can be used for building combinations. Here, we used common levels to combine models (prognostic scores) [64]:
  - Combination level:Mainly focuses on possible ways of combining the risk outputs of prognostic methods in an ensemble based on the information obtained from a single method.
  - Classifier level:Different machine learning techniques can be employed.
  - Feature level: Use patients' prognostic factors for machine learning.

4. Create a subset $D_c$of conflict group of patients from DP as rows represent patientP_i and columns representPS_mdata. The last column is the actual molecular responseY_i:
$$Y_{i=}\begin{cases} 0 & notachivingmolcularresponse \\ 1 & achiveingmolcularresponse \end{cases}$$

5. Divide$D_c$into training and testing datasets, and apply internal validation on training data.

6. Create the combined prognostic model $M_c$specifically for$D_c$to classifyY_i using different combined functions that maximize the accuracy of the classification.

7. Evaluate the combined method by the combined function (classifier) by comparing the actual outcome Y_i withthe method outcome$Y_i M_c$, and measure the model performance.

8. Generate the evaluation dataset$E$, where the rows are all possible models$M_c$that are built from different combined methods and the columns contain the performance measurements accuracy, sensitivity, specificity, PPV, NPV, G-mean and F-score.

9. Rank the model $M_c$based on selected criteria.

10. Select the best models$M_c$that resolve the conflict.

11. Evaluate the selected model $M_c$onall $P_i$in DP.

12. Compare the performance of$M_c$with the previous methods using testing data.

13. Perform external validation on an unseen dataset.

2.5.2 Dealing with a New Source of Clinical Data (Predictive Model)

The schema for the predictive model is illustrated in Figure 5. Algorithms are often used to select a relevant subset of input features (in our problem, a subset of predictive factors that will deliver a highly predictive model)[59, 65].
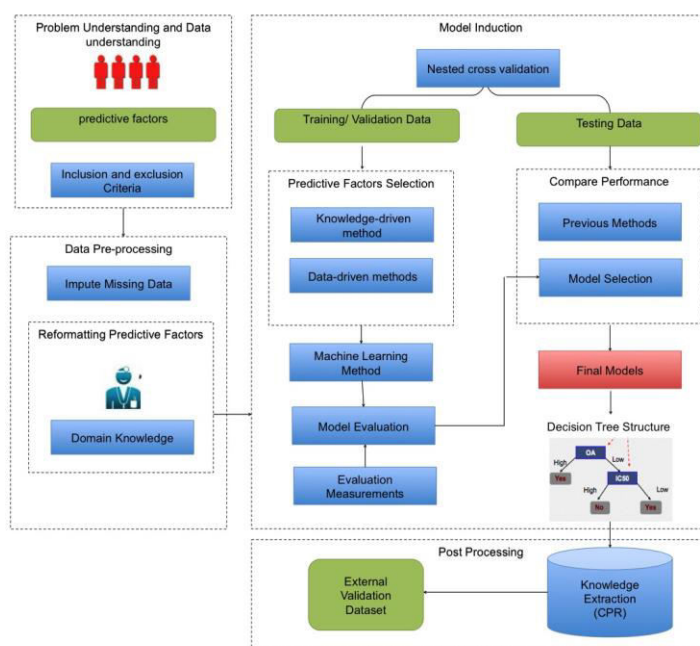
Figure 5. The schema of the predictive model

This is also very important in the context of healthcare costs, where fewer input factors imply fewer diagnostic tests to obtain the relevant predictive factors [66]. We also need to extract the relations between the most related predictive factors and try to understand whether there exist clinical rules for prediction. We divided the feature selection process into two main types:

- A knowledge-driven method for feature selection, such as existing medical literature that has published predictive factors as an informative feature or clinical expert judgment on molecular factors associated with predicting molecular response, known as manual feature selection [67].
- Data-driven methods for feature selection, known as automatic feature selection. We used the wrapper approach [58], where all subsets of the features are evaluated using a given machine learning approach.

Model induction can be briefly described as follows:

1. Apply predictive factor Pf_z selection on dataset D to identify which predictive factors are important to be used as an input vector.
2. Create sample x from D where rows represent patient P_i, and the columns are predictive factor Pf_z . The last column is the actual molecular response Y_i:
3. $Y_{i=} \begin{cases} 0 & not achiving molcular response \\ 1 & achiveing molcular response \end{cases}$
4. Divide x into training and testing datasets, and apply internal validation to the training data.
5. Create the predictive model M_x specifically for x to predict Y_i using the learning function that maximizes the prediction accuracy.
6. Evaluate the predictive factor subset by the learning function (classifier) by comparing the actual outcome Y_i with the predicted outcomes $Y_i M_x$ and measure the model performance.
7. Generate the evaluation dataset E, where the rows are all possible models M_c that are built from different combined methods and the columns contain the performance measurements accuracy, sensitivity, specificity, PPV, NPV, G-mean and F-score.

15

8. Rank the model based on selected criteria.
9. Select the best models.
10. Compare the performance of M_x with the previous methods using testing data.
11. Perform external validation on an unseen dataset.
12. Visualize the structure of the model.
13. Generate the clinical prediction rules.

Clinicalprediction rules from data: in the clinical decision process, clinical prediction rules play an important role after increasing the acceptance of evidence-based medicine. This is because of the advantages that can be gained by using clinical prediction rules. First, these rules can support and guide clinicians in patient casesinvolving complex decisions. Second, these rules,which represent expert knowledge, can be shared with clinicians in primary care. Finally, these rules can be used to train and guide new experts.The decision tree structure represents extracted production rules [68]. We used the training set of patients from which a decision tree was generated to create a structure. The path from the root of the tree to the leaf node was used to establish the conditions (IF parts), whereas every leaf of a decision tree corresponded to a response (THEN parts). The decision rules R were in the form, IF Pf_1 is C_1 AND Pf_2 is C_2, THEN Response is Y, where Pf_1 and Pf_2 are the predictive factors, C_1 and C_2 are the subcategories that belong to Pf_1 and Pf_2, respectively, and Y, is the class (molecular response achieved or not achieved).

2.6 Evolving Approaches

In a personalized medicine support system, there are three possible schemas for the evaluation of new knowledge: confirmation, contradiction and contribution [24]. In confirmation, no changes need to be made to the existing knowledge base. In contribution, the new knowledge is used for constructing predictive models and updating clinical prediction rules (predictive modeling steps in this study). In contradiction, the new knowledge conflicts with the old knowledge, and combined methods are used to resolve the conflict in knowledge (prognostic model steps in this study).

The intelligent techniques used in the proposed framework are selected to generalize the predictive and prognosis performance. The predictive factor wrapper approach evaluates all possible combinations of predictive factors. Therefore, new predictive factors will be equally tested in the construction of predictive models. Prognostic modeling evaluates different combination levels. Therefore, a new level of knowledge, either prognostic factors, prognostic scores or prognostic model, can be combined and evaluated. Nested cross validation is a suitable validation method for evaluation, as internal validation can be used to evaluate the new knowledge and inform the decision of whether to enhance the performance, refine the modeling or ignore the new knowledge in case of no further improvement.The following suggestions are given as solutions and will be applied in a personalized medicine support system automatically if the internal validation of the selected model does not outperform a previous method in the unseen dataset (testing data).An evolving system can change in the following ways:

- Data pre-processing phase in predictive model: the expert range of the predictive factors of conflict cases can be changed, and this should be confirmed by domain experts.
- Predictive factor selection approaches: predictive factors or prognostic factors can be updated or added.
- Combined method: different combination levels can be selected.

3. PROPOSED APPLICATIONS

Personalized medicine support systems will provide great benefits in the future as molecular data analysis improves. This type of framework based on individual data will become a basic tool used by healthcare professionalsfor selecting personalized treatments. Clinicians will be able to allocate risk profiles and predict molecular responses to selected molecular targeted therapy, thereby improving treatment selection and avoiding disease progression. The treatment of many types of cancer can be explored using this framework. Table 1 shows some of the cancers that can be treated with multiple molecular targeted therapies as well as the available studies for prognostic and predictive factors.

Table 1 of the cancers that can be treated with multiple molecular targeted therapies as well as the available prognostic and predictive factors to use for the same

Type of Cancer	FDA-approved molecular targeted therapy [69]	Prediction and Prognosis
Breast cancer	Ado-trastuzumabemtansine (Kadcyla), Everolimus (Afinitor), Lapatinib (Tykerb), Palbociclib (Ibrance), Pertuzumab (Perjeta), Trastuzumab (Herceptin)	[70]
Lung cancer	Afatinib (Gilotrif), Alectinib (Alecensa), Atezolizumab (Tecentriq), Ceritinib (Zykadia),Erlotinib (Tarceva), Gefitinib (Iressa), Necitumumab (Portrazza), Nivolumab (Opdivo), Osimertinib (Tagrisso), Pembrolizumab (Keytruda), Ramucirumab (Cyramza)	[71]
Melanoma	Aldesleukin (Proleukin),Dabrafenib (Tafinlar),Ipilimumab (Yervoy), Nivolumab (Opdivo), Pembrolizumab (Keytruda), Trametinib (Mekinist), Vemurafenib (Zelboraf)	[72]
Chronic myeloid leukemia	Imatinib (Gleevec), Nilotinib (Tasigna), Dasatinib (Sprycel), Ponatinib (Iclusig), Bosutinib (Bosulif)	[73 , 74]
Chronic lymphocytic leukemia	Alemtuzumab (Campath), Idelalisib (Zydelig), Obinutuzumab (Gazyva), Ofatumumab (Arzerra, HuMax-CD20), Rituximab (Rituxan, Mabthera), Venetoclax (Venclexta)	[75]

4. CONCLUSIONS

The framework of current learning methodologies basically includesthe essential concept of dynamic changes in the input structure of models based on domain knowledge and discoveringthe relation of importance predictive factors with patient responses. In our framework, we conduct precise modeling based on prior knowledge with interpretable and consistent meanings,and we focus on developing models with high predictive and prognostic performance.

Owing to space and time limitations, the implementation and some computational approaches have not been included in this paper. The initial future work is to implement a framework for real cancer patient data that stores molecular responses to molecular targeted therapy. Important areas for extending our proposed approach are as follows:(1) optimization techniques should be tested for model selection;(2) we trained models to maximize accuracy, but algorithms could also be developed to maximize the G-mean or F-score performance, especially in the imbalance dataset;(3) the use of a wider range of combined methods such as decision templates and the Dempster-Shafer method[76] should be enabled; and(4) online adaptive models that dynamically react to a new piece of knowledge should be developed.

In this paper, we have described a novel framework for a personalized medicine support system. A personalized medicine support system is a hybrid knowledge-based system that integrates multiple models to provide automated selection tools for individual treatment in one integrated dynamic environment. This system has been proposed to meet the requirements of clinicians and to provide a treatment guide. This system can work on oncology-based data using a data mining process for modeling; this approach can be implemented for treating other types of cancer aswell.

REFERENCES

[1] I. S. Kohane, "The twin questions of personalized medicine: who are you and whom do you most resemble?," BioMed Central, vol. 1, pp. 4.1-4.3, 2009.

[2] M. Verma, "Personalized Medicine and Cancer," Journal of Personalized Medicine, vol. 2, pp. 1-14, 2012.

[3] M. Müller, "Individualized medicine," in Clinical Pharmacology: Current Topics and Case Studies, ed: Springer, 2010.

[4] D. L. Mitchell, "Successful Implementation of Personalized Medicine: The Value, Challenges , and Effect on Patient Care," Master of Science in Health Care Administration, the Faculty of Utica College, ProQuest, 2013.

[5] M. Dietel and C. Sers, "Personalized medicine and development of targeted therapies: the upcoming challenge for diagnostic molecular pathology. A review," Virchows Archiv, vol. 448, pp. 744-755, 2006.

[6] M. Sioud and Ø. Melien, "Treatment Options and Individualized Medicine," Target Discovery and Validation Reviews and Protocols, vol. 361, pp. 327-340, 2007.

[7] H. Michael and M. Crawford, "Individualized Medicine by Biomarkers," JAMA, vol. 302, pp. 49-57, 2009.

[8] B. Shastry, "Pharmacogenetics and the concept of individualized medicine," The Pharmacogenomics Journal vol. 6, pp. 16-21, 2006.

[9] J. Corander, T. Aittokallio, S. Ripatti, and S. Kaski, "The rocky road to personalized medicine: computational and statistical challenges," Personalized Medicine: Future Medicine part of Future Science Group, vol. 9, pp. 109–114, 2012.

[10] L. M. Fagan and E. H. Shortliffe, "The Future of Computer Applications in Biomedicine," in Biomedical Informatics Computer Applications in Health Care and Biomedicine, 3 ed USA: Springer, 2006, pp. 829-847.

[11] R. T. Ng and J. Pei, "Introduction to the special issue on data mining for health informatics," SIGKDD Explor. Newsl., vol. 9, pp. 1-2, 2007.

[12] M. Kurosaki, "Pretreatment prediction of response to peginterferon plus ribavirin therapy in genotype 1 chronic hepatitis C using data mining analysis," Journal of gastroenterology (0944-1174), vol. 46, p. 104, 2011.

[13] A. D. Revell, D. Wang, M. A. Boyd, S. Emery, A. L. Pozniak, F. D. Wolf, et al., "The development of an expert system to predict virological response to HIV therapy as part of an online treatment support tool," AIDS, vol. 25, pp. 1855–1863 2011.

[14] C. E. Pedreira, L. Macrini, M. G. Land, and E. S. Costa, "New Decision Support Tool for Treatment Intensity Choice in Childhood Acute Lymphoblastic Leukemia," IEEE Transaction on Information Technology in Biomedicine, vol. 13, pp. 284-290, 2009.

[15] H. Ying, F. Lin, R. D. MacArthur, J. A. Cohn, D. C. Barth-Jones, H. Ye, et al., "A Fuzzy Discrete Event System Approach to Determining Optimal HIV/AIDS Treatment Regimens," IEEE Transactions on Information Technology in Biomedicine vol. 10, p. 663, 2006.

[16] M. A. Musen, Y. Shahar, and E. H. Shortliffe, "Clinical Decision-Support Systems," in Biomedical Informatics Computer Applications in Health Care and Biomedicine, 3rd ed USA: Springer, 2006, pp. 698-736.

[17] M. Ball, C. Weaver, and J. Kiel, "Overview of Clinical Decision Support Systems," in Healthcare Information Management Systems, Third, Ed., ed New York: Springer-Verlag, pp. 463-477.

[18] F. Alonso, L. Martínez, A. Pérez, and J. P. Valente, "Cooperation between expert knowledge and data mining discovered knowledge: Lessons learned," Expert Systems with Applications, vol. 39, pp. 7524–7535, 2012.

[19] M. A. Musen, B. Middleton, and R. A. Greenes, "Clinical Decision-Support Systems," in Biomedical Informatics: Computer Applications in Health Care and Biomedicine, E. H. Shortliffe and J. J. Cimino, Eds., ed London: Springer London, 2014, pp. 643-674.

[20] L. Aleksovska-Stojkovska and S. Loskovska, "Clinical Decision Support Systems: Medical knowledge acquisition and representation methods," in Electro/Information Technology (EIT), 2010 IEEE International Conference on, 2010, pp. 1-6.

[21] E. S. Berner, "Clinical Decision Support Systems: State of the Art," Agency for Healthcare Research and Quality, 2009.

[22] D. Levinson, "Information, computers, and clinical practice," Jama, vol. 249, pp. 607-9, 1983.

[23] Y. Hu, N. Kasabov, and W. Liang, "Personalized Information Modeling for Personalized Medicine," Springer Handbook of Bio-/Neuroinformatics, pp. 533-553, 2014.

[24] M. Kwiatkowska and A. S. Atkins, "Integrating knowledge-driven and data-driven approaches for the derivation of clinical prediction rules," in Machine Learning and Applications, 2005. Proceedings. Fourth International Conference on, 2005, p. 6 pp.

[25] N. Esfandiari, M. R. Babavalian, A.-M. E. Moghadam, and V. K. Tabar, "Knowledge discovery in medicine: Current issue and future trend," Expert Systems with Applications, vol. 41, pp. 4434-4463, 7// 2014.

[26] N. Kasabov and D. Filev, "Evolving Intelligent Systems: Methods, Learning, & Applications," International Symposium on Evolving Fuzzy Systems, IEEE, 2006.

[27] S. T. C. Wong, "Coping with Conflict in Cooperative Knowledge-Based Systems," IEEE Tranaction on Sytems, man, And Cybernetics-Part Asystems and Humans, vol. 27, 1997.

[28] M. K. Yusof and A. Azlan, "Comparative Study of Techniques in Reducing Inconsistent Data," International Journal of Database Theory and Application, vol. 5, pp. 37-47, 2012.

[29] A. H. Bond and L. Gasser, "An analysis of problems and research in DAI," in Readings in Distributed ArtiJicial Intelligence, ed, 1988.

[30] W. T. Adrian, A. Lig̦eza, and G. J. Nalepa, "Inconsistency Handling in Collaborative Knowledge Management," in Federated Conference on Computer Science and Information Systems, 2013, pp. 1221-1226.

[31] G. Ke, B. Fu, M. De, and K. Hipel, "A hierarchical multiple criteria model for eliciting relative preferences in conflict situations," Journal of Systems Science and Systems Engineering, vol. 21, pp. 56-76, 2012/03/01 2012.

[32] F. J. J. d. Santos and H. d. A. Camargo, "Decision Support Systems in Multicriteria Groups: an Approach Based on Fuzzy Rules," IEEE, 2010.

[33] B. S. Ahn and S. H. Choi, "Conflict resolution in a knowledge-based system using multiple attribute decision-making," 2009.

[34] M. Noor-E-Alam, T. F. Lipi, M. Ahsan Akhtar Hasin, and A. M. M. S. Ullah, "Algorithms for fuzzy multi expert multi criteria decision making (ME-MCDM)," Knowledge-Based Systems, vol. 24, pp. 367-377, 2011.

[35] H. Junming, S. Chong, L. Shuang, and W. Wanshan, "Fuzzy Case-Based Reasoning for Conflict Resolution in Collaborative Design," in Computing, Communication, Control, and Management, 2008. CCCM '08. ISECS International Colloquium on, 2008, pp. 233-237.

[36] G. Qi, W. Liu, and D. Bell, "Measuring conflict and agreement between two prioritized knowledge bases in possibilistic logic," Fuzzy Sets and Systems, vol. 161, pp. 1906-1925, 7/16/ 2010.

[37] K. Mu, K. Wang, and L. Wen, "Approaches to measuring inconsistency for stratified knowledge bases," International Journal of Approximate Reasoning, vol. 55, pp. 529-556, 1// 2014.

[38] K. Mu, Z. Jin, W. Liu, D. Zowghi, and B. Wei, "Measuring the significance of inconsistency in the Viewpoints framework," Science of Computer Programming, vol. 78, pp. 1572-1599, 9/1/ 2013.

[39] J. Grant and A. Hunter, "Measuring inconsistency in knowledgebases," Journal of Intelligent Information Systems, vol. 27, pp. 159-184, 2006.

[40] A. Hunter and S. Konieczny, "Approaches to Measuring Inconsistent Information," in Inconsistency Tolerance, L. Bertossi, A. Hunter, and T. Schaub, Eds., ed Berlin, Heidelberg: Springer Berlin Heidelberg, 2005, pp. 191-236.

[41] N. Kasabov, "Soft Computing Methods for Global, Local and Personalized Modeling and Applications in Bioinformatics," Soft ComputingBased Modeling in Intel. Systems, vol. 196, pp. 1-18, 2009.

[42] K. M. Khalil, M. Abdel-Aziz, T. T. Nazmy, and A.-B. M. Salem, "Intelligent Techniques for Resolving Conflicts of Knowledge in Multi-Agent Decision Support Systems," presented at the Sixth International Conference on Intelligence Computing and Information Systems, Cairo, Egypt, 2014.

[43] . Australian Clinical Trials. Available: https://www.australianclinicaltrials.gov.au/what-clinical-trial/phases-clinical-trials

[44] . National Institutes of Health. https://www.nih.gov.

[45] S. Saussele and M. Pfirrmann, "Clinical Trials in Chronic Myeloid Leukemia," Curr Hematol Malig Rep pp. 109–115, 2012.

[46] B. M. Welch and K. Kawamoto, "The Need for Clinical Decision Support Integrated with the Electronic Health Record for the Clinical Application of Whole Genome Sequencing Information," J. Pers. Med., vol. 3, 2013.

[47] J. C. Weiss , S. Natarajan, P. L. Peissig, C. A. McCarty, and D. Page, "Machine learning for personalized medicine: predicting primary myocardial infarction from electronic health records," AI Magazine, vol. 33, p. 33, 2012.

[48] C. L. Overby, "A Clinical Decision Support Model for Incorporating Pharmacogenomics Knowledge Into Electronic Health Records for Drug Therapy Individualization: A Microcosm of Personalized Medicine," Doctor of Philosophy, Division of Biomedical and Health Informatics, University of Washington, ProQuest, 2011.

[49] I. Kouris, C. Tsirmpas, S. G. Mougiakakou, D. Iliopoulou, and D. Koutsouris, "E-Health towards ecumenical framework for personalized medicine via Decision Support System," in Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE, 2010, pp. 2881-2885.

[50] B. R. Hirsch and A. P. Abernethy, "Structured Decision-Making: Using Personalized Medicine to Improve the Value of Cancer Care," Journal of Personalized Medicine, vol. 3, pp. 1-13, 2013.

[51] S. Mehta, A. Shelling, A. Muthukaruppan, A. Lasham, C. Blenkiron, G. Laking, et al., "Predictive and prognostic molecular markers for cancer medicine," Therapeutic Advances in Medical Oncology, vol. 2, pp. 125-148, 2010.

[52] R. Alsolami, S. J. L. Knight, and A. Schuh, "Clinical application of targeted and genome-wide technologies: can we predict treatment responses in chronic lymphocytic leukemia?," Personalized Medicine, vol. 10, pp. 361-376, 2013/06/01 2013.

[53] A. Lopez-Chavez, A. Thomas, A. Rajan, M. Raffeld, B. Morrow, R. Kelly, et al., "Molecular profiling and targeted therapy for advanced thoracic malignancies: a biomarker-derived, multiarm, multihistology phase II basket trial," J Clin Oncol, vol. 33, pp. 1000-7, Mar 20 2015.

[54] J. M. Llovet and J. Bruix, "Molecular targeted therapies in hepatocellular carcinoma," Hepatology, vol. 48, pp. 1312-27, Oct 2008.

[55] P. P. Piccaluga, G. Martinelli, and M. Baccarani, "Advances in the treatment for haematological malignancies," Expert Opin Pharmacother, vol. 7, pp. 721-32, Apr 2006.

[56] G. M. Grimshaw, A. Szczepura, M. Hulten, F. MacDonald, N. C. Nevin, F. Sutton, et al., "Evaluation of molecular tests for prenatal diagnosis of chromosome abnormalities," Health Technol Assess, vol. 7, pp. 1-146, 2003.

[57] . National Comprehensive Cancer Network Guideline Inc.

[58] A. G. Karegowda, M. A. Jayaram, and A. S. Manjunath, "Feature Subset Selection Problem using Wrapper Approach in Supervised Learning," International Journal of Computer Applications, vol. 1, pp. 13–17, 2010.

[59] P. Perner, "Improving the accuracy of decision tree induction by feature preselection " Applied Artificial Intelligence: An International Journal, vol. 15, 2001.

[60] L. Breiman, J. Friedman, R. Olshen, and C. Stone, Classification and regression trees: Belmont, Calif : Wadsworth International Group, 1984.

[61] J. Lasserre, S. Arnold, M. Vingron, P. Reinke, and C. Hinrichs, Predicting the outcome of renal transplantation vol. 19, 2012.

[62] N. S. Russell and A. C. Begg, "Editorial radiotherapy and oncology 2002: predictive assays for normal tissue damage," Radiotherapy and Oncology, vol. 64, pp. 125-129, 8// 2002.

[63] D. Steinley, "Standardizing Variables in K-means Clustering," in Classification, Clustering, and Data Mining Applications: Proceedings of the Meeting of the International Federation of Classification Societies (IFCS), Illinois Institute of Technology, Chicago, 15–18 July 2004, D. Banks, F. R. McMorris, P. Arabie, and W. Gaul, Eds., ed Berlin, Heidelberg: Springer Berlin Heidelberg, 2004, pp. 53-60.

[64] L. I. Kuncheva, Combining Pattern Classifiers Methods and Algorithms: John Wiley & Sons,, 2004.

[65] R. C. Prati, "Combining feature ranking algorithms through rank aggregation," in Neural Networks (IJCNN), The 2012 International Joint Conference on, 2012, pp. 1-8.

[66] M. Qian, I. Nahum-Shani, and S. A. Murphy, "Dynamic Treatment Regimes," in Modern Clinical Trial Analysis, W. Tang and X. Tu, Eds., ed New York: Springer, 2013, pp. 127-148.

[67] T.-H. Cheng, C.-P. Wei, and V. S. Tseng, "Feature Selection for Medical Data Mining: Comparisons of Expert Judgment and Automatic Approaches," in Proceedings of the 19th IEEE Symposium on Computer-Based Medical Systems (CBMS'06), 2006.

[68] J. R. Quinlan, "Generating production rules from decision trees," in Proceeding IJCAI'87 Proceedings of the 10th international joint conference on Artificial intelligence, USA, 1987, pp. 304-307

[69] R. Abramson. (2016). Overview of Targeted Therapies for Cancer. Available: https://www.mycancergenome.org/content/molecular-medicine/overview-of-targeted-therapies-for-cancer/

[70] T. Foukakis and J. Bergh. (2016). Prognostic and predictive factors in early, nonmetastatic breast cancer. Available: https://www.uptodate.com/contents/prognostic-and-predictive-factors-in-early-nonmetastatic-breast-cancer

[71] M. Paesmans, "Prognostic and predictive factors for lung cancer," Breathe, vol. 9, pp. 112-121, 2012.

[72] D. Weinstein, J. Leininger, C. Hamby, and B. Safai, "Diagnostic and Prognostic Biomarkers in Melanoma," The Journal of Clinical and Aesthetic Dermatology, vol. 7, pp. 13-24, 2014.

[73] K. Sweet, L. Zhang, and J. Pinilla-Ibarz, "Biomarkers for determining the prognosis in chronic myelogenous leukemia," Journal of Hematology & Oncology, vol. 6, 2013.

[74] D. L. White and T. P. Hughes, "Predicting the response of CML patients to tyrosine kinase inhibitor therapy," Current Hematologic Malignancy Reports, vol. 6, pp. 88-95, 2011.

[75] D. Mertens and S. Stilgenbauer, "Prognostic and Predictive Factors in Patients With Chronic Lymphocytic Leukemia: Relevant in the Era of Novel Treatment Approaches?," Journal of Clinical Oncology, vol. 32, pp. 869-872, 2014.

[76] A. Altmann, M. Rosen-Zvi, M. Prosperi, E. Aharoni, H. Neuvirth, E. Schülter, et al., "Comparison of Classifier Fusion Methods for Predicting Response to Anti HIV-1 Therapy," PLoS ONE, vol. 3, 2008.