Journal of Medical Systems (2021) 45: 10
https://doi.org/10.1007/s10916-020-01679-3

TRANSACTIONAL PROCESSING SYSTEMS

An Integrated Platform for Skin Cancer Heterogenous and Multilayered Data Management

Ilias Maglogiannis[1] · Georgia Kontogianni[1,2,3] · Olga Papadodima[2] · Haralampos Karanikas[4] · Antonis Billiris[5] · Aristotelis Chatziioannou[2,3,6]

Received: 11 April 2019 / Accepted: 23 November 2020 / Published online: 6 January 2021

Abstract

Electronic health record (EHR) systems improve health care services by allowing the combination of health data with clinical decision support features and clinical image analyses. This study presents a modular and distributed platform that is able to integrate and accommodate heterogeneous, multidimensional (omics, histological images and clinical) data for the multi-angled portrayal and management of skin cancer patients. The proposed design offers a layered analytical framework as an expansion of current EHR systems, which can integrate high-volume molecular -omics data, imaging data, as well as relevant clinical observations. We present a case study in the field of dermatology, where we attempt to combine the multilayered information for the early detection and characterization of melanoma. The specific architecture aspires to lower the barrier for the introduction of personalized therapeutic approaches, towards precision medicine. The paper describes the technical issues of implementation, along with an initial evaluation of the system and discussion.

Keywords Clinical decision support tools · Composite biomarkers · Dermoscopy · Melanoma · Next generation sequencing

Introduction

Modern Holistic Electronic Healthcare systems aim to improve the provided health care services by offering services that combine health data with other features such as clinical decision support and data analytics to medical professionals. This combination may lead to major health care savings, reduce medical errors, upgrade the quality health services and affect several health Key Performance Indicators (KPIs). Such sophisticated Clinical Decision Support tools (CDS) provide for instance patient-specific advice on medication intakes based on previous treatments, they calculate distance from therapeutic goals, and offer other clinical recommendations. Promising next generation developments will include prioritizing clinical actions that have maximum benefit to a given patient at the point of care and developing effective methods to communicate CDS information to patients to better incorporate patient preferences in care decisions [1].

Skin cancer is considered one of the most frequent types of cancer. One in every three cancers diagnosed is a skin cancer, and according to the Skin Cancer Foundation Statistics, three out of ten Caucasians will develop skin cancer during their lifetime. Amongst the most common skin cancers is basal cell carcinoma that causes significant inconvenience to a person's life due to high recurrence but is rarely deadly since it generally does not metastasize. In contrast, although melanoma is an infrequent type of skin cancer, it is considered among the most lethal forms of cancer. Other types of skin cancer rarely spread to other parts of the body, but melanomas are considered an aggressive type of cancer, with high metastatic potential. Skin cancer incidence has increased the past few decades and specifically in the case of cutaneous melanoma, incidence rates in Caucasian populations have risen faster than any other malignant entity over the last 30 years. Melanoma incidence

✉ Ilias Maglogiannis
imaglo@unipi.gr

[1] Department of Digital Systems, University of Piraeus, 126 Grigoriou Lambraki, 18534 Piraeus, Greece

[2] National Hellenic Research Foundation, 48 Vassileos Constantinou Ave, 11635 Athens, Greece

[3] Center of Systems Biology, Biomedical Research Foundation of the Academy of Athens, 4 Soranou Ephessiou, 11527 Athens, Greece

[4] University of Thessaly, Papasiopoulou 2-4, Lamia, Greece

[5] Datamed S.A., Grammou 71, Maroussi, Greece

[6] e-NIOS Applications Private Company, 17671 Kallithea, Greece

has been increasing since the mid-60s and is predicted to keep increasing in most fair skinned populations [2, 3].

As far as melanoma phenotypic classification is concerned, experts divide the disease into several subgroups [4]. The four most common clinical subtypes include superficial spreading melanoma, lentigo malignant melanoma, nodular melanoma and acral lentiginous melanoma. Other rare variants include desmoplastic melanoma and nevoid melanoma. Towards the molecular characterization of melanoma, next generation sequencing (NGS) technologies are a valuable tool and have been exploited in a number of studies [5–8], comparing sequencing data from melanoma tissue and a matched normal control in order to identify somatic mutations, radically reshaping our understanding regarding the high complexity of the genomic landscape of this disease. Towards this end, the largest genomic analysis of cutaneous melanoma from the Cancer Genome Atlas (TCGA) Network proposed four major subtypes: BRAF-mutated, RAS-mutated, NF1-mutated and triple-wild-type [9]. Characterizing a patient's mutation profile can lead to the administration of tailored drugs, aiding in the attainment of personalized precision medicine [10]. Goal of this study is to design and implement a layered analytical framework as an expansion of current EHR systems, which will be able to integrate EHR data, high-volume molecular -omics data, dermoscopic data of skin lesions, and other relevant clinical observations. The motivation of our research is to integrate all levels of melanoma related health data and to incorporate all this information in a Knowledge Base and a Clinical Decision Support System for Dermatology. The overall concept of the platform is illustrated in Fig. 1.

In this paper, we describe the related work and background information in Section II, while in Section III details about the technical issues of design and implementation are provided. Section IV includes an initial evaluation of the implemented system and finally, Section V concludes the paper.

Related work and background information

Integrated electronic health record (EHR) systems and clinical decision support systems

Improved clinical decision making via the meaningful use of EHRs is of paramount importance. Consistent use of an EHR does not always ensure successful exploitation for improving quality of care. Ongoing efforts to encourage meaningful use and adoption should focus on advance tools that will upgrade the functionalities of an EHR system in the direction of clinical decision support. In the context of this work, we have incorporated wider (concerning diversity) data sources including clinical examination, image dermoscopy and molecular data in conjunction with state-of-the-art analysis techniques. The followed approach relies on early adopters that had begun to integrate clinical decision support systems (CDSS) [10, 11]. These types of systems are supporting a variety of topics and they are designed to assist the medical personnel in all stages starting from initial consultation to diagnosis and follow up.

Our main focus was data integration at EHR level, which is about how to combine data from a large variety of heterogeneous sources into meaningful and valuable information. Data from different systems need to be integrated technically as well as semantically. In order to achieve semantic interoperability in the healthcare domain numerous standardization efforts are in place in order to define common information models (or data elements) such that all systems can operate with data on the same level. In our case this was done either in

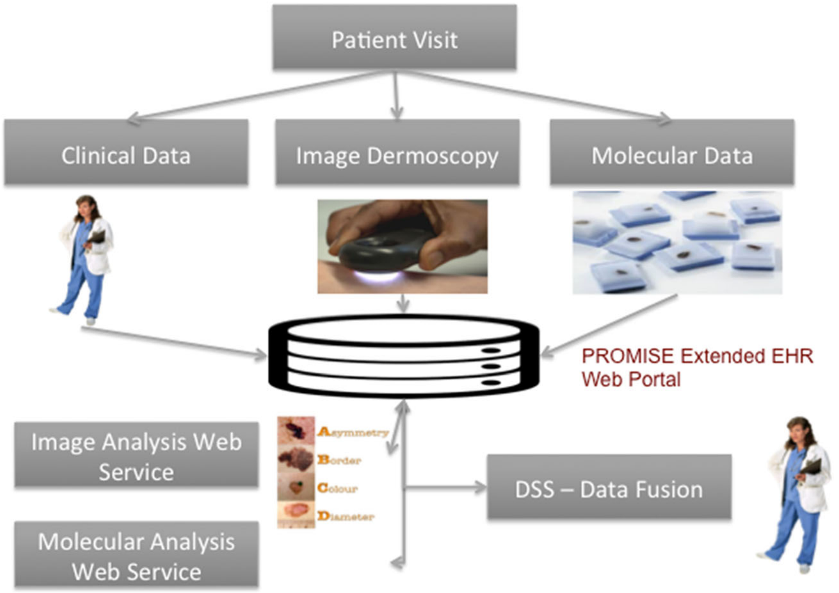


Fig. 1 The concept of the integrated platform for Melanoma data Management

the primary data level or the secondary data and indicators level. Collecting data directly from the subjects interested in, is called primary data. Data gathered by someone else is called secondary. Semantic interoperability can be defined as the ability of two or more computer systems to exchange information in such a way that the "meaning" of that information can be automatically interpreted by the receiving system accurately enough to produce useful results to the end users of both systems [12]. Many standardization efforts focus on EHRs in order to facilitate integration of electronic health data accumulating in healthcare facilities and the most important are:

- openEHR – an open standard for health data based on a complete separation between software and clinical models, thus ensuring universal interoperability.
- Health Level Seven (HL7) – a set of standards for transfer of clinical and administrative data between software applications.
- Integrating the Healthcare Enterprise (IHE) – an initiative by healthcare professionals and industry to improve the way computer systems in healthcare share information. IHE promotes the coordinated use of established standards such as DICOM and HL7 to address specific clinical needs in support of optimal patient care.

There are two main paradigms for the semantic interoperability between EHR-systems: Messaging Paradigm and the Two-Level-Modelling Paradigm. The SemanticHealthNet EU-project addressed this problem quite deeply by involving the major stakeholders to solve this problem, such as the EN13606 Association, International Health Terminology Standards Development Organization (IHTSDO) and the WHO. In our work data exchange, is performed using the CDA (Clinical Document Architecture) format, which is one of the most widely used standards of HL7. The CDAs exchanged is formatted in XML format and the structure is modeled following the EPSOS standard, so as to ensure maximum interoperability with other future information health systems. Additionally, a Service oriented architecture (SOA) has been adopted to solve the interoperability of the involved heterogeneous subsystems, something which is widely adopted for distributed EHR systems [13]. This architecture plays a key role in the integration of heterogeneous systems by means of services that represent different system functionality, independent of the underlying platforms or programing languages, and interacts via message exchanges. Web services also play a critical role in systems' interoperability. Web services technology is defined as a systematic and extensible framework for application's interactions that is built on top of existing web protocols. These protocols are based on XML [14] and include: Web Services Description Language (WSDL) to describe the service interfaces, Simple Object Access Protocol (SOAP) for communication between web services and client applications, and Universal Description, Discovery, and Integration (UDDI), to facilitate locating and using web services on a network [15].

Harness the power of health data with the use of Big Data technologies is a rather new technology for the e-Health domain. There are still a lot of issues that need to be resolved before efficient health data analytics can be performed. One of the most important issues is the binding of data (to patients, situations, sensor devices). Since data are coming from different sources, this metadata must be bound to patient IDs. The wide variation of data models and data warehouses with their own data binding is also a current problem. A new approach, called late-binding architectures, delays data binding until the proper time and context, and retains the collected data its original, undiluted value. In these new Big Data architectures, the repository for structured, unstructured and semi-structured data in its original format is generally called "Data Lake". Having healthcare as one of the popular use cases for Big Data and analytics, we are recently observing several implementations of these architectures and concepts in healthcare [16, 17]. We recognize as a challenge that processing EHR data only for data exchange is not enough for exploiting the power of existing data. The semantics of EHR data should be linked with the data coming from other sources also. Our solution for EHR integration does not incorporate the Big Data architecture, since it was not needed up to now, but it is something that we are planning for the near future. Accumulating data in data warehouses and steaming data from different sources challenge the existing architectures and existing approaches fall short to meet the requirements in this Big Data world. For example, the challenges now are linking a diagnosis from a patient's EHR with the data coming from wearables on that patient and performing data analytics to help physicians for predictive medicine or even to help the patient take actions against bad conditions.

Analytics is to make sense of your data and uncover meaningful trends [18]. Analytics is defined as the method of "logical analysis". A method of logical analysis is commonly performed using algorithms. This applied logic produces a model in which the parts are related with statistical relevance. Often, analytics is future-oriented, predicting relations, whereas analysis is associated with what is or has been. Integrating Data Analytics in operational Healthcare Information systems requires [19] the use of full range and huge amount of heterogeneous information including electronic medical records, images and sensors that we refer as big data. The extraordinary potential to the exploitation of these amounts of valuable information by using a combination of machine learning and data mining tools will improve patient care process and patient life quality [20, 21]. According to [22], current health care systems under development or in production are lacking the potential benefits of big data analytics [23].

When it comes to the exchange and analysis of skin specific or general healthcare data, there are several challenges. Interoperability of healthcare data exists to some extent, but the proliferation of common data element models does not help to solve the interoperability problem. Current solutions are not capable of processing data from different sources in different formats. The challenge is to achieve a form of automatic data format harmonization such that data from new applications and devices can easily be added independent of the origin of the data. With respect to data analytics, we must solve important limitations related to the trustworthiness and heterogeneity of the sources. It is highly desirable to clean data in advance of analyzing it and using it to make life-or-death decisions. Moreover, it is important to develop optimized methods for dealing with data quality issues before implementing classification, regression or clustering algorithms. Medical image analytics is in its infancy. There is great need for developing a platform and new algorithms that allow to use (big) data from other sources in the analysis and interpretation of images. When it comes to using the collected and analyzed data, current decision support systems lack the flexibility required to accept massive amounts of data from heterogeneous sources on the one hand and to use these data for personalized decision support on the other. Related to this is the need to enable medical professionals to define and tune the decision support rules themselves in human language without the intervention of programmers or an information specialist. The above principles are adopted during the implementation of the integrated system as described in Section III.

Digital image analysis in dermoscopy

As already mentioned, modern electronic healthcare systems are being upgraded to efficiently manage large medical image databases. Furthermore, image analysis tools are being incorporated, so as to automate the extraction of meaningful features from medical images and assist diagnosis [24]. Specifically, in the field of digital dermoscopy, there is a plethora of image features that are found in the literature, which provide useful information for image assisted diagnosis of skin cancer lesions. The most prominent are the ones based on the ABCD rule, and the ABCDE rule (Asymmetry, Border, Color, Diameter, Evolution), the Menzies scale and the Seven-point scale [25–27]. The major advantage of computer extracted features is reproducibility, making dermoscopy image evaluation more objective, while human interpretation of image characteristics can be subjective [28–30]. Therefore, computer based expert systems have been introduced as alternatives and adjuncts to the naked-eye expert assessment. Very detailed and comprehensive reviews of such systems and the corresponding technologies may be found in [31, 32]. Most of the proposed systems aim at the detection of malignant melanoma at early stages versus dysplastic or common nevus using images in the visible spectrum [34–37]. Infrared or ultraviolet illumination (in situ or in vivo) using appropriate multispectral cameras are also used in [36–38]. Microscopy (or epiluminence microscopy) setups are found in the works of [39, 40] and digital videomicroscopy in [41].

The features that are used for computer based skin lesion analysis are mostly those associated with color in various color representation spaces (RGB, HIS, CIELab), e.g., color values in [33, 35, 42] and Colorbin (i.e., the percentage of the lesion colored in foreground pixels) [42]. Some of the approaches used feature combination in more than one color spaces for achieving better results, e.g., both RGB and IHS in [34, 39, 43, 44], or RBG and colors peculiar to malignant melanomas in [45]. The light intensity features are also used in works like [38]. Asymmetry and border based features are also used extensively e.g., [40–42], while features based on specific differential structures are rare. Some papers [46, 47] rely also on the whole ABCD rule for lesion characterization, while in some cases shape and color features, like Area and Elevation, are calculated manually by dermatologists [42].

Common classification methods are the rule-based ones, e.g., [34, 37–39, 41, 46, 48]. Advanced machine learning techniques such as neural networks [49] and support vector machines are also presented in works like [33, 35, 42, 47, 50–52], while the k-nearest neighborhood classification scheme is applied in [40]. Evidence Theory (upper and lower probabilities induced by multivalued mapping) based on the concept of lower and upper bounds for a set of compatible probability distributions is used in [53]. The reported accuracies and results concerning the works presented in the literature prove that image based automated classification of lesions and melanoma in particular may work. More detailed descriptions and additional results regarding the various methods used in existing dermoscopy analysis systems are presented in [41]. The most popular of them are adopted for the implementation of Dermoscopy Imaging Analysis Service, which is an important part of the integrated system, as described in Section III.

Integration of molecular analysis and composite biomarkers

In addition to medical imaging, the onset and constant advancement of molecular technologies has enabled the parallel, high-throughput process of millions of sequence reads, thus ushering a new era with numerous, novel applications in basic, applied and clinical research. An important class of molecular technologies encompasses gene expression technologies (microarrays analysis or RNA sequencing). The meticulous monitoring of the haplome, namely the universe of variations concerning the genome of a species, if related to the pathology under investigation, allows the exploration of the impact of those variations in the gene expression between different phenotypic groups. More importantly it allows the extrapolation of profiling patterns of genomic sequences (whole genome, whole exome (WES), or targeted sequencing of a gene panel) with classification ability for the different phenotypic classes of a disease/pathology. In the context of this work, we aimed to integrate the different levels of molecular data, so as to produce a robust diagnostic signature for the classification of melanoma.

Coalescing diverse levels of information improves the total knowledge on a problem and promotes its resolution [54]. Based on this, diagnosis should be based on the correct integration of molecular, histological and clinical features, so as to become more accurate. Previous analyses were able to achieve better performance on given tasks, through combination of heterogeneous data [54, 55], or by building multi-marker models for accurate classification of melanoma [56–58].

Better understanding of the etiological aspects and mechanisms of cancer development are vital to improve survival rate and prevention. Given this perspective, recent studies have shown an improved performance, when combining transcriptomics with gene regulatory data in ovarian cancer [59]. Efficient predictive biomarkers from multiple approaches or different levels of analyses support optimal characterization of the tumor under investigation. Gene signature strategies are tested extensively for their potential to transform clinical practice i.e. to support immunotherapy-based, management of cancer-patients [60].

Our previous work [61, 62] has led to the discovery of 32 critical genes, whose expression offers key information on melanoma manifestation. Here, we intend to extend this knowledge to mutational data. Ultimate aim is to produce a robust diagnostic gene signature that will allow the classification of the patients and at the same time aid in the context of personalized medicine.

Platform architecture and integration

Overall architecture

A distributed, architecture for the EHR system implementation, was chosen as illustrated in Fig. 2. The architecture will work well in distributed environments in which health facilities has each own local EHR it satisfies the interoperability standards of the EHR. The system also has a set of distributed prediction services (knowledge bases, external classifiers) handling the various levels of data. Each external service is specialized in a specific domain (i.e. prediction based on image metadata, prediction based on the integration of molecular and image data) and the framework achieves cooperation and integration of these services and the central system in order the final user to face a unified experience.

The data layer in addition to traditional EHRs, includes dermatological specific data as shown in Fig. 3 and the related image and molecular metadata. Internationally standards for the exchange of medical data were followed such as the CDA (Clinical Document Architecture) format, which is one of the most widely used standards of HL7. The CDAs exchanged is formatted in XML format and the structure is following EPSOS standard [63], so as to ensure maximum interoperability with other information health systems. Coding Standards also included ICD-10 for diagnosis, ICPC2 for diagnosis on primary care and ATC5 for Medicines-Drugs). In the business layer in order to support prediction based on the classifier built with the molecular and image data, a web service exists to link to the external classifiers. The user level refers to all the public and private Healthcare units (Hospitals, Health Centers, Doctors and others healthcare providers/ professionals), which are interconnected with the Central Agency Management System of the patients' dermatological electronic health records (EHR). Through these interfaces the need for access to the Dermatological EHR of the patient is served.

The user via the web portal is able to input relevant data on the EHR's database and on demand the system will acquire predictions based on the classifiers built and trained constantly. Based on the electronic prediction information, EHR could relate various heterogeneous data with the actual clinical and other patient data usually stored in EHR systems. Moreover, the integration of EHR with the Dermatology decision support system, by making individual treatment information available regardless of time and location, can increase patient's right to know, improve the ability to manage diseases, and alleviate the asymmetry between the medical staff and the patient. Our primary goal was to enhance the EHR with the ability to maintain relevant to the skin cancer cases information and to provide prediction based on historical data.

Dermoscopy imaging analysis service

As already mentioned in dermoscopy image analysis, feature design is based on the so-called ABCD-rule of dermatology. ABCD rule, which constitutes the basis for a diagnosis by a dermatologist represents the Asymmetry, Border structure, variegated Color, and the Differential Structures of the skin lesion. The feature extraction is performed by measurements on the pixels that represent a segmented object allowing non-visible features to be computed. In this context, the implemented image analysis web service includes three (3) types of imaging features, which were calculated as follows: Border Features covering the A and B parts of the ABCD-rule of dermatology, Color Features which correspond to the A and C rules, and Textural Features which are corresponding to D rules [64]. The feature extraction procedure resulted in a total of thirty-one (31) features. The feature extraction

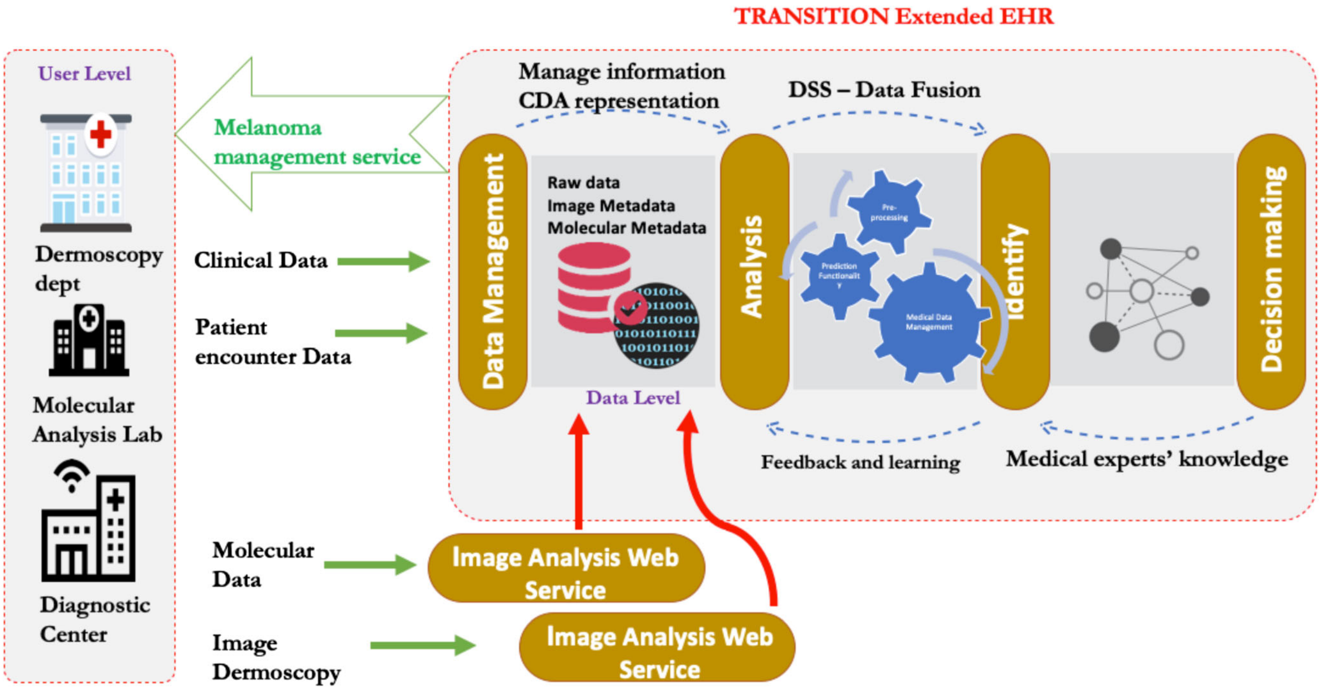


Fig. 2 System architecture of the integrated platform for Melanoma data Management

methodology and the feature assessment are presented in previous works [62, 65] correspondingly. The relevant preprocessing for all features is described in [62]. The specific image analysis functionality has been integrated as a service in the integrated platform. The input of the service is the dermoscopy image stored in the database and the output is the 31 calculated ABCD image features.

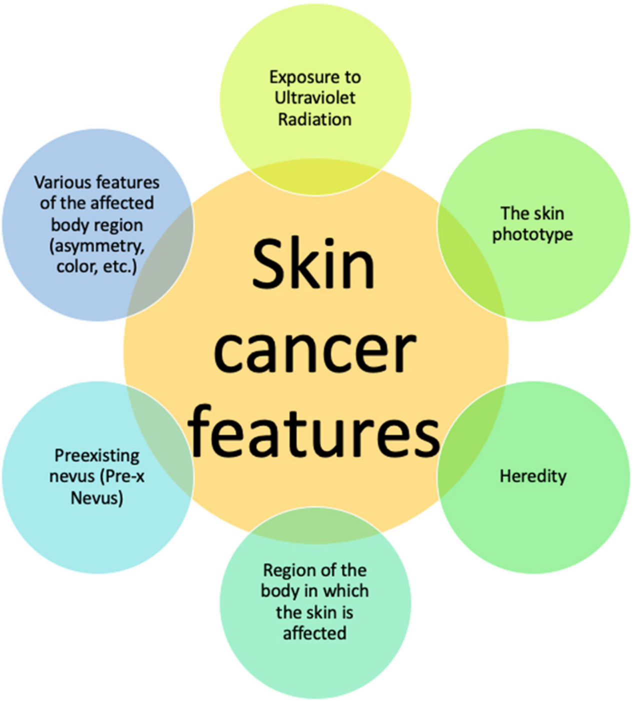

Skin cancer features
Exposure to Ultraviolet Radiation
The skin phototype
Heredity
Region of the body in which the skin is affected
Preexisting nevus (Pre-x Nevus)
Various features of the affected body region (asymmetry, color, etc.)

Fig. 3 PROMISE expanded EHR includes features especially for the skin cancer prediction case

Molecular data analysis service

Molecular Data Analysis concerns raw next generation sequencing (NGS) data derived from exome sequencing of melanoma tissue and matched healthy control. The framework of analysis of NGS data has been previously presented by our team [66]. A pilot analysis was performed including eight patients [67]. The outcome of this analysis is a list of significantly mutated genes for melanoma. As it was previously stated this list is constantly updated, as new patients are added in our database.

Next, we sought to build a classifier exploiting the mutational data that were produced. Since the number of patients analyzed in this work was limited, we added samples from TCGA database through cBioPortal [68, 69]. As healthy state (non-melanoma) we used mutational data from dysplastic nevi that were acquired through similar experimental procedure [70]. On the molecular level, this state holds a considerably lower mutational load compared to melanoma, and few mutated genes in total, 232 genes, as opposed to the 1586 genes, found in our case [67]. For feature selection, we reduced the list of mutated genes, to a total of 51 genes (molecular signature, see Supplementary Table 1), by distinguishing the 'driver' mutations, i.e. mutations with high impact on the product, using PolyPhen2 [71], and then prioritizing them according to their centrality (genes taking part in numerous distinct mechanisms are ranked higher), using BioInfoMiner

[72]. Through BioInfoMiner, we aimed to distinguish putatively causative genes, assuming that genes with implication in diverse cross-talking biological processes -reflecting genes with a central role in cellular physiology- could be promising candidates and have a great impact in the manifestation of the disease. BioInfoMiner exploits semantic information to detect and rank genes based on their centrality, as described in different databases (e.g. Gene Ontology [73, 74], Reactome [75], etc.). The entropy-based information gain ratio, a metric that expresses the amount of information contained in a given attribute, was used to assess the list of molecular features. The top genes with the highest information gain ratio and their corresponding ranking by BioInfoMiner are presented in Table 1.

The samples (samples of dysplastic nevi and melanomas) were separated under two labels, dysplastic nevus (*represented by DNS*) and melanoma (*represented by MEL*) and each sample is attributed a 51-dimensional binary vector showing whether the corresponding gene contains a mutation or not. To deal with unbalanced classes, the SMOTE [76] algorithm was utilized to generate synthetic data for the DNS label. This data assortment is presented in Fig. 4. Due to the binary type of the classification problem, the *Random Forests (RF)* algorithm [77] was selected, as an appropriate and effective methodology. Additional classification algorithms were examined, generally showing equivalent outcome, due to the evident discrepancy of the two classes (see Fig. 5). RF implementations are often more parametrizable than similar tree-based algorithms (like *Decision Trees*) and this permitted an exhaustive grid search for fine-tuning of classification parameters. Also, RF is a recursive algorithm, an asset that prevents being trapped in a subset of solutions and so all contingencies are included, with the appropriate statistical weight. Here, the R programming language was used [78], and packages caret [79], DMwR [80] and pROC [81].

Table 1 List of important mutated genes sorted in descending order according to their information gain ratio from the 51 gene signature used as input for the molecular classifier, together with their corresponding topological centrality score calculated by BioInfoMiner, describing the association of each gene with the given number of distinct biological modules

Genes	Information gain ratio	Topological centrality score
ANK3	0.334544	36
RELN	0.284221	2
GRIN2A	0.280225	46
SCN5A	0.268364	32
FLT1	0.252773	22
COL3A1	0.23731	4
KALRN	0.23731	8
CFTR	0.233444	48
ROBO2	0.229572	44
LAMA2	0.22569	37
NRXN1	0.22569	45
DMD	0.221794	13
EPHA7	0.221794	20
CELSR1	0.221794	40
ANGPT1	0.21788	7
CACNA1C	0.213943	35
PTPRO	0.193696	47
KDR	0.189475	11
PPP1R9A	0.185172	25
NR1H4	0.185172	42
CARMIL1	0.185172	51
LRRK2	0.180769	6
PKP2	0.180769	14
POSTN	0.176247	28

The best performance was reported for the RF classifier with the following parameters:

- 122 samples for training, 120 for testing
- 51 predictors
- 2 classes: 'DNS', 'MEL'
- No pre-processing
- Resampling: Cross-Validated (10 fold, repeated 3 times)
- mtry = 26

As a criterion for the cross validation performance, the receiver operating characteristic (ROC) curve was used, which controls the sensitivity with respect to the specificity [82]. The area under the curve (AUC) of the plot gives an unbiased estimation of the classifier's performance at each round. The classifier performed very well, reaching a mean accuracy of 0.93. This result

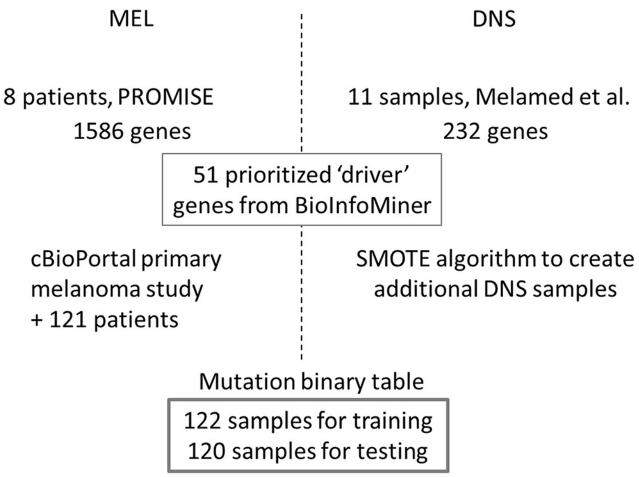


Fig. 4 Data assortment for classification

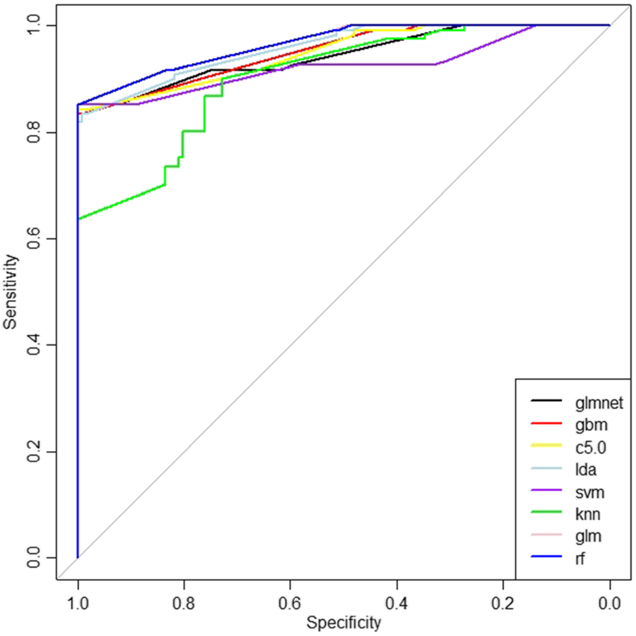


Fig. 5 Results for the Molecular Classifier. ROC curve for rf-Random Forests, glmnet-Guassian linear model, gbm-Stochastic Gradient Boosting, c50-Decision Trees C5.0, lda-Linear Discriminant Analysis, svm-Support Vector Machines, knn-k Nearest Neighbours, glm-Logistic Regression

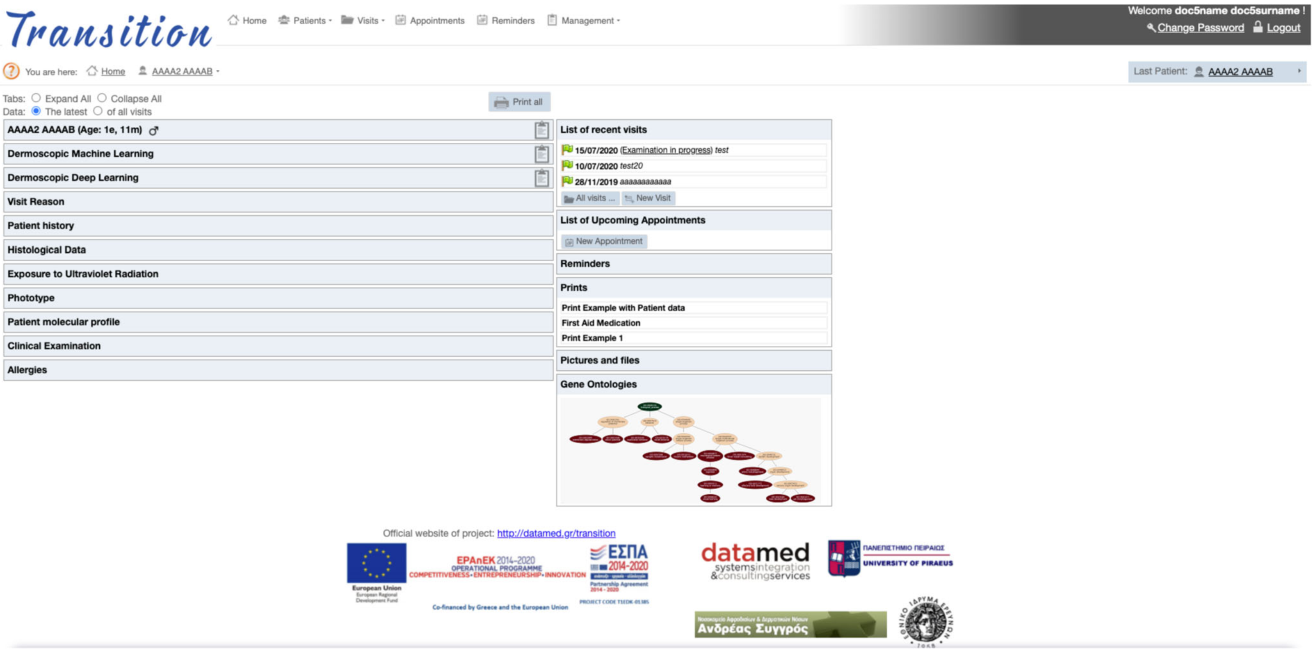
justifies the utilization of this classifier as a model for class prediction (melanoma vs. dysplastic nevus) of unknown samples of mutation data. The final model was stored in order to be used for the implementation of the web service, which will accept the data and perform the predictions.

The system in practice case study evaluation and results

The integrated platform for skin cancer related healthcare data management in practice

The implemented system is capable of integrating seamlessly multiple sources of heterogenous and bulky data, concerning the results of either high-throughput molecular analysis (DNA microarrays or NGS) or dermoscopic (epiluminescence dermoscopy) examination. The system's font-end is a Web based SPA (Single Page Application) as depicted in Fig. 6. Our goal was to extend the patient's EHR by integrating pertinent clinical, molecular and imaging data in order to support clinical decision-making and prediction. This is accomplished through the integration of multi-layered dermatological data and its interoperability with the Expert System for Melanoma Recognition, which correlates molecular and imaging data to obtain a prediction. The comprehensive analysis of clinical, NGS and imaging data, as well as other available high efficiency data, enables the construction of a network model for the melanoma disease including different phenotypic categories as well as different organization levels. The EHR expansion is called PROMISE as it was funded by the project: "Personalization of melanoma therapeutic management through the fusion of systems biology and intelligent data mining methodologies".

The medical record in each tab contains a different category of medical information. In dermatological view tab there are two basic options. In model user marks the area in human body that appears to have a problem. Color and size of this


Transition
Home
Patients ▾
Visits ▾
Appointments
Reminders
Management ▾
Welcome doc5name doc5surname !
Change Password
Logout
Print all
All visits ...
New Visit
New Appointment
ΕΡΑνΕΚ 2014-2020 OPERATIONAL PROGRAMME COMPETITIVENESS•ENTREPRENEURSHIP•INNOVATION
European Union European Regional Development Fund
ΕΣΠΑ 2014-2020
Partnership Agreement 2014 - 2020
PROJECT CODE T1EDK-01385
Co-financed by Greece and the European Union
datamed systemsintegration &consultingservices
ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ UNIVERSITY OF PIRAEUS
Ανδρέας Συγγρός
ΕΘΝΙΚΟ ΙΔΡΥΜΑ ΕΡΕΥΝΩΝ

You are here: Home AAAA2 AAAAB ▾

Last Patient: AAAA2 AAAAB

Tabs: ○ Expand All ○ Collapse All
Data: ● The latest ○ of all visits

AAAA2 AAAAB (Age: 1e, 11m) ♂
Dermoscopic Machine Learning
Dermoscopic Deep Learning
Visit Reason
Patient history
Histological Data
Exposure to Ultraviolet Radiation
Phototype
Patient molecular profile
Clinical Examination
Allergies

List of recent visits
15/07/2020 (Examination in progress) *test*
10/07/2020 *test20*
28/11/2019 *aaaaaaaaaaaa*

List of Upcoming Appointments

Reminders

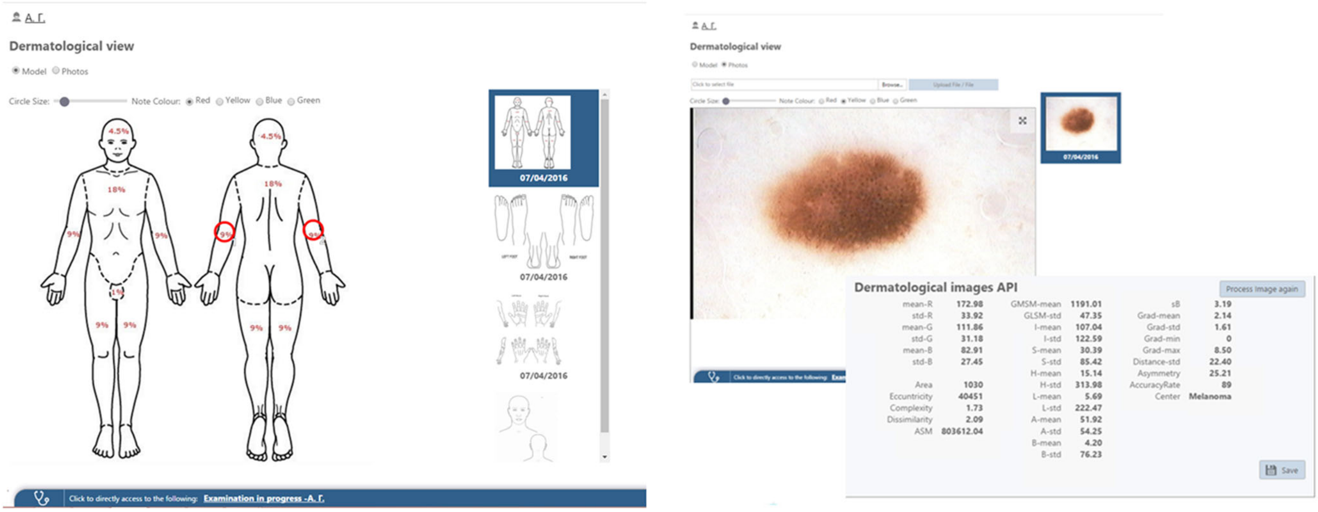
Prints
Print Example with Patient data
First Aid Medication
Print Example 1

Pictures and files

Gene Ontologies

Official website of project: http://datamed.gr/transition

Fig. 6 Main tabs of Melanoma Information Management Tool



A.Γ.

Dermatological view

● Model ○ Photos

Circle Size: Note Colour: ● Red ○ Yellow ○ Blue ○ Green

Click to directly access to the following: Examination in progress -A. Γ.

A.Γ.

Dermatological view

○ Model ● Photos

Click to select file Browse... Upload File / File

Circle Size: Note Colour: ○ Red ● Yellow ○ Blue ○ Green

07/04/2016

Dermatological images API

Process Image again

mean-R	172.98	GMSM-mean	1191.01	sB	3.19
std-R	33.92	GLSM-std	47.35	Grad-mean	2.14
mean-G	111.86	I-mean	107.04	Grad-std	1.61
std-G	31.18	I-std	122.59	Grad-min	0
mean-B	82.91	S-mean	30.39	Grad-max	8.50
std-B	27.45	S-std	85.42	Distance-std	22.40
		H-mean	15.14	Asymmetry	25.21
Area	1030	H-std	313.98	AccuracyRate	89
Eccuntricity	40451	L-mean	5.69	Center	Melanoma
Complexity	1.73	L-std	222.47		
Dissimilarity	2.09	A-mean	51.92		
ASM	803612.04	A-std	54.25		
		B-mean	4.20		
		B-std	76.23		

Save

Click to directly access to the following: Exam

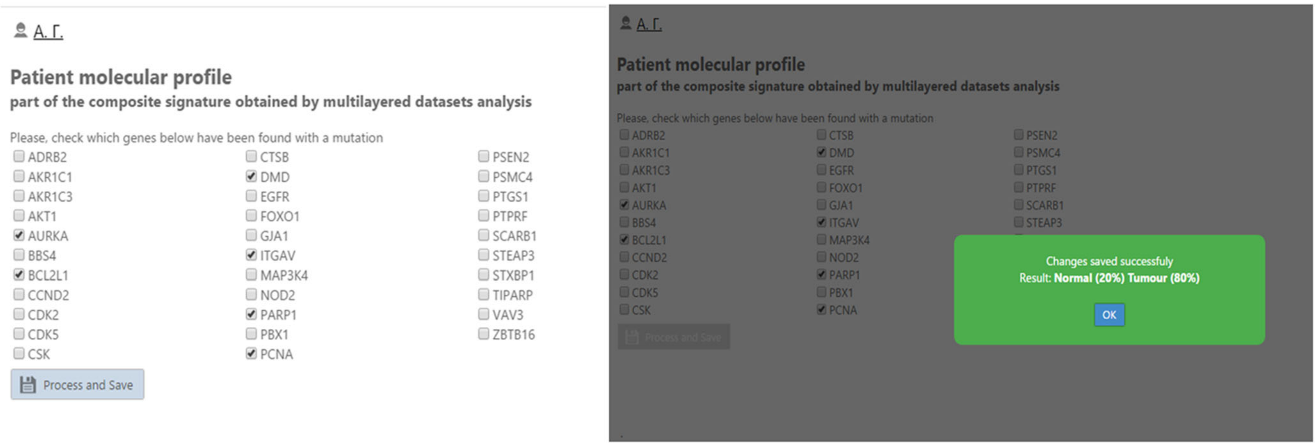
Fig. 7 Tab category on dermatological view and image analysis view

circle are configurable by user. In "Photos" option, user is able to upload a dermoscopy image of the concerning case. The Dermoscopy Image Analysis Service is then called, and image is analyzed in its features as shown as output (see Fig. 7).

In Histological Data tab user fills out patient's histological data. All data are shown in different dropdown menus with predefined values. In clinical examination tab user fills out patient's clinical examination data. In Patient Molecular Profile tab, the genes of the molecular signature are selected by checkboxes. User crosses those genes of the patient that bear a mutation. By choosing "Process and save" molecular analysis classifier is called. A pop-up message is shown with classifier's decision. Result shows two percentages for normal and tumor (see Fig. 8).

Qualitative and quantitative results

In order to validate the proof of concept of the proposed design we created a fused dataset containing molecular and clinical data for the melanoma case. More specifically, a synthetic dataset was constructed to incorporate images from different nevi (dysplastic or melanomas) together with molecular measurements which are encountered in the same stages, using the imaging and WES data that were available, described in sections III.B and III.C. The nature of the molecular features allows for this 'random' integration, due to the small number of mutations, especially in the dysplastic nevus class. In total, tests were performed on three (3) different datasets i) the molecular dataset of 51 features (section III.C), ii) the imaging dataset of 31 features and iii) the integrated dataset of 51 + 31 features. The parameters used for each RF classifier are similar (section III.C), apart from the number of predictors used, that equals the number of features. Performance metrics obtained by classification modules (DNS vs. MEL) support that integrated features perform best, regarding the discrimination between malignant and benign sample classes, and constitute to an improved classifier, compared to the molecular and imaging classifiers. From the



A. Γ.

Patient molecular profile

part of the composite signature obtained by multilayered datasets analysis

Please, check which genes below have been found with a mutation

□ ADRB2
□ AKR1C1
□ AKR1C3
□ AKT1
☑ AURKA
□ BBS4
☑ BCL2L1
□ CCND2
□ CDK2
□ CDK5
□ CSK
□ CTSB
☑ DMD
□ EGFR
□ FOXO1
□ GJA1
☑ ITGAV
□ MAP3K4
□ NOD2
☑ PARP1
□ PBX1
☑ PCNA
□ PSEN2
□ PSMC4
□ PTGS1
□ PTPRF
□ SCARB1
□ STEAP3
□ STXBP1
□ TIPARP
□ VAV3
□ ZBTB16

Process and Save

A. Γ.

Patient molecular profile

part of the composite signature obtained by multilayered datasets analysis

Please, check which genes below have been found with a mutation

□ ADRB2
□ AKR1C1
□ AKR1C3
□ AKT1
☑ AURKA
□ BBS4
☑ BCL2L1
□ CCND2
□ CDK2
□ CDK5
□ CSK
□ CTSB
☑ DMD
□ EGFR
□ FOXO1
□ GJA1
☑ ITGAV
□ MAP3K4
□ NOD2
☑ PARP1
□ PBX1
☑ PCNA
□ PSEN2
□ PSMC4
□ PTGS1
□ PTPRF
□ SCARB1
□ STEAP3

Changes saved successfuly

Result: **Normal (20%) Tumour (80%)**

OK

Process and Save

Fig. 8 Molecular Profile view, a typical snapshot

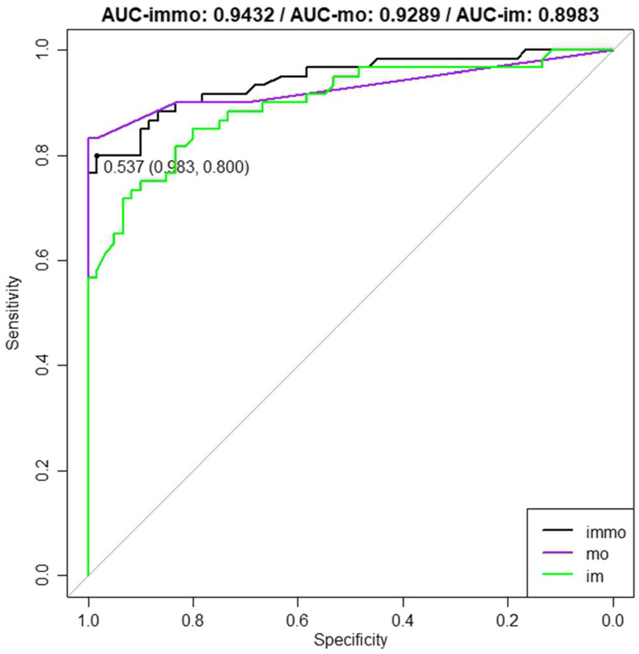


Fig. 9 ROC curves for the 3 Random Forests classifiers, immo-integrated features (82) classifier, mo-molecular features (51) classifier, im-imaging features (31) classifier. The integrated feature classifier performs best, with a mean AUC of 0.9432

statistical perspective, the use of synthetic data is more conservative when the number of replicates is large. Essentially, it is the closest and more plausible approach to be adopted for simulation purposes. The corresponding results in the form of ROC curves are illustrated in Fig. 9.

The entropy-based information gain (IG) and the information gain ratio (GR) were measured for the set of features from the integrated dataset. Four extra datasets were created containing the top 10 and top 20 features per measurement, to evaluate the classification accuracy. Performance metrics obtained by classification modules (DNS vs. MEL) support that all integrated feature datasets perform equally good, as shown in Fig. 10.

Success of classification depends mainly on feature selection. Based on this, the total list of features (82) achieves high accuracy, improved only by the top features taken through gain ratio. Still, cutting the number of features does not necessarily improve this system, since the performance was equal to begin with, plus extra information is considered for the therapeutic approach, when required.

User friendliness and acceptance evaluation

A number of 10 physicians (equal representation of men and women) was recruited for assessing the implemented tool. A questionnaire was built on the criteria of the Core Objectives for Eligible Professionals - Stage 2 of the EHR Incentive Programs (2018) and on SUS questionnaire, which was proposed by John Brook in 1986 and can provide a high-level subjective view of usability [83, 84]. At first stage we presented the platform and its features to each respondent, who were then given time to use the system, before the evaluation. In addition, we conducted interviews with the physicians in order to qualitative assess the platform. The outcome of this analysis was that the platform has been designed according to well-known requirements for physicians' usability. Usability testing with the SUS method has provided an overall positive evaluation of the system's design. Qualitative research gave more insights on the needs and preferences of physicians and has proposed some improvements in regard to user interface design, functionalities that appeal to users and some ways to make the users' interaction with the system more convenient for them. Our findings have illustrated the importance of incorporating prediction features in EHR, as well as features that can make the system more attractive, in order to motivate physicians to use them on a regular basis.

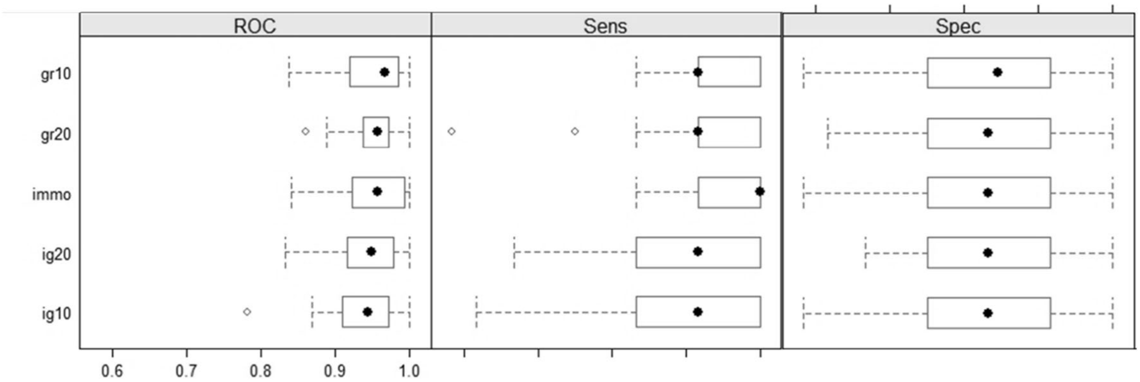


Fig. 10 Results for the Integrated RF Classifiers, ROC, Sensitivity and Specificity for immo-82 features, gr10-gain ratio top-10, gr20-gain ratio top-20, ig10-information gain top-10 and ig20- information gain top-20 features

Conclusions

Concluding, in this work we propose an advanced EHR system, which is specially designed for the field of dermatology and aims to integrate various levels and sources of heterogenous data linked with skin diseases and particularly skin cancer. The described system is based on a Web based distributed EHR architecture utilizing web services, while it adopts established best practices, regarding data collection and management, integration and integrity, from operationally tested clinical decision systems. Our goal is to provide a proof of concept system that expands the analytical, administrative and interpretive capabilities of existing EHR systems, adopting machine leaning and decision support functionalities. More specifically, the presented system proposes a paradigm, which through the massive integration of multi-layered, heterogeneous data, depicting phenotypic aspects of the disease manifestation and the parallel processing of those streams, independently but also in relation with each other, will produce appropriate sets of composite biomarkers that ultimately assist and accelerate medical diagnosis and patient therapeutic management. In this context, such a system could enable the introduction of personalized approaches in the therapeutic course. The distributed and modular architecture allows the accommodation of additional experimental protocols, either in the area of molecular analysis (i.e. single cell genomics, tissue heterogeneity studies, time-course analysis) or in supporting more dermoscopy imaging modalities (i.e. confocal or multispectral dermoscopy). Each new module can be integrated in the system in the form of a new web service. It remains now as future work to include additional sources of skin cancer related data and to perform more experiments in order to prove the diagnostic value of the presented integration schemes and further validate the design of the implemented system.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10916-020-01679-3.

Acknowledgements This work has been supported by TRANSITION project (MIS 5031295) which is implemented under the Action “Reinforcement of the Research and Innovation Infrastructure”, funded by the Operational Programme“ Competitiveness, Entrepreneurship and Innovation” (NSRF 2014-2020) and co-financed by Greece and the European Union (European Regional Development Fund). G. Kontogianni’s research is co-financed by Greece and the European Union (European Social Fund- ESF) through the Operational Programme «Human Resources Development, Education and Lifelong Learning» in the context of the project “Reinforcement of Postdoctoral Researchers - 2nd Cycle” (MIS-5033021), implemented by the State Scholarships Foundation (IKΥ).

Funding This work has been supported by TRANSITION project (MIS 5031295) which is implemented under the Action “Reinforcement of the Research and Innovation Infrastructure”, funded by the Operational Programme“ Competitiveness, Entrepreneurship and Innovation” (NSRF 2014–2020) and co-financed by Greece and the European Union (European Regional Development Fund).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval Since all datasets included in this article were either from public databases or from previously published studies additional approval by an ethics committee was unnecessary. All human samples were acquired in the context of a prior project entitled 12CHN-204 PROMISE (Bilateral Greece-China Research Program of the Hellenic General Secretariat of Research and Technology and the Chinese Ministry of Research and Technology entitled “Personalization of melanoma therapeutic management through the fusion of systems biology and intelligent data mining methodologies-PROMISE”, sponsored by the Program “Competitiveness and Entrepreneurship”, Priority Health of the Peripheral Entrepreneurial Program of Attiki), under the strict conformity to the rules of the call. All procedures performed in studies involving human participants were in accordance with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.”

References

1. P. J. O’Connor, J. R. Desai, J. C. Butler, E. O. Kharbanda, and J. M. Sperl-Hillen, “Current status and future prospects for electronic point-of-care clinical decision support in diabetes care,” *Current diabetes reports*, vol. 13, no. 2, pp. 172–176, 2013.
2. E. De Vries, V. De Poll-Franse, W. Louwman, F. De Gruijl, and J. Coebergh, “Predictions of skin cancer incidence in the Netherlands up to 2015,” *British Journal of Dermatology*, vol. 152, no. 3, pp. 481–488, 2005.
3. G. P. Guy *et al.*, “Vital signs: melanoma incidence and mortality trends and projections - United States, 1982–2030,” *MMWR Morb. Mortal. Wkly. Rep.*, vol. 64, no. 21, pp. 591–596, 2015.
4. A. M. Bailey *et al.*, “Implementation of biomarker-driven cancer therapy: existing tools and remaining gaps,” *Discovery medicine*, vol. 17, no. 92, p. 101, 2014.
5. K. Dutton-Regester and N. K. Hayward, “Reviewing the somatic genetics of melanoma: from current to future analytical approaches, ” *Pigment cell & melanoma research*, vol. 25, no. 2, pp. 144–54, 2012, https://doi.org/10.1111/j.1755-148X.2012.00975.x.
6. X. Wei *et al.*, “Exome sequencing identifies GRIN2A as frequently mutated in melanoma,” *Nature genetics*, vol. 43, no. 5, pp. 442–446, 2011.
7. M. Krauthammer *et al.*, “Exome sequencing identifies recurrent somatic RAC1 mutations in melanoma,” *Nature genetics*, vol. 44, no. 9, pp. 1006–14, 2012, https://doi.org/10.1038/ng.2359.
8. E. Hodis *et al.*, “A landscape of driver mutations in melanoma,” *Cell*, vol. 150, no. 2, pp. 251–63, 2012, https://doi.org/10.1016/j.cell.2012.06.024.
9. “Genomic Classification of Cutaneous Melanoma,” *Cell*, vol. 161, no. 7, pp. 1681–96, 2015, https://doi.org/10.1016/j.cell.2015.05.044.
10. C. Castaneda *et al.*, “Clinical decision support systems for improving diagnostic accuracy and achieving precision medicine,” *Journal of clinical bioinformatics*, vol. 5, no. 1, p. 4, 2015.
11. L. Kuhn *et al.*, “Planning for action: the impact of an asthma action plan decision support tool integrated into an electronic health record (EHR) at a large health care system,” *The Journal of the American Board of Family Medicine*, vol. 28, no. 3, pp. 382–393, 2015.
12. W. Ceusters and B. Smith, “Semantic Interoperability in Healthcare State of the Art in the US,” *New York State Center of Excellence in Bioinformatics and Life Sciences Ontology Research Group*, pp. 1–33, 2010.

13. C. Hahn, S. Jacobi, and D. Raber, “Enhancing the interoperability between multiagent systems and service-oriented architectures through a model-driven approach,” 2010, vol. 2, pp. 415–422.
14. World Wide Web Consortium, *2012*. 2012.
15. J. Bacon and K. Moody, “Toward open, secure, widely distributed services,” *Communications of the ACM*, vol. 45, no. 6, pp. 59–64, 2002.
16. H. Catalyst, *Late-Binding Data Warehouse, Health Catalyst.*.
17. *Diving in: Navigating a data lake for predictive care Patient Data Intelligence fo Next-Generation Care Delivery*.
18. M. M., “The Difference Between Data, Analytics, and Insights,” *Localytics*, Dec. 2016. http://info.localytics.com/blog/difference-between-data-analytics-insights.
19. A. T. Janke, D. L. Overbeek, K. E. Kocher, and P. D. Levy, “Exploring the potential of predictive analytics and big data in emergency care,” *Annals of emergency medicine*, vol. 67, no. 2, pp. 227–236, 2016.
20. A. Holzinger and I. Jurisica, “Knowledge discovery and data mining in biomedical informatics: The future is in integrative, interactive machine learning solutions,” in *Interactive knowledge discovery and data mining in biomedical informatics*, Springer, 2014, pp. 1–18.
21. M. Šprogar, M. Lenič, and S. Alayon, “Evolution in Medical Decision Making,” *Journal of Medical Systems*, vol. 26, no. 5, pp. 479–489, 2002, https://doi.org/10.1023/A:1016413418549.
22. F. Wang, L. S. Docherty, K. J. Turner, M. Kolberg, and E. H. Magill, “Services and policies for care at home,” 2006, pp. 1–10.
23. N. T. Issa, S. W. Byers, and S. Dakshanamurthy, “Big data: the next frontier for innovation in therapeutics and healthcare,” *Expert review of clinical pharmacology*, vol. 7, no. 3, pp. 293–298, 2014.
24. T. Goudas and I. Maglogiannis, “An advanced image analysis tool for the quantification and characterization of breast cancer in microscopy images,” *Journal of medical systems*, vol. 39, no. 3, p. 31, 2015.
25. G. Argenziano, G. Fabbrocini, P. Carli, V. De Giorgi, E. Sammarco, and M. Delfino, “Epiluminescence microscopy for the diagnosis of doubtful melanocytic skin lesions. Comparison of the ABCD rule of dermatoscopy and a new 7-point checklist based on pattern analysis,” *Archives of dermatology*, vol. 134, no. 12, pp. 1563–70, 1998.
26. G. Betta, G. Di Leo, G. Fabbrocini, A. Paolillo, and M. Scalvenzi, “Automated Application of the ‘7-point checklist’ Diagnosis Method for Skin Lesions: Estimation of Chromatic and Shape Parameters,” 2005, vol. 3, pp. 1818–1822.
27. M. Ogorzałek, L. Nowak, G. Surowka, and A. Alekseenko, “Melanoma in the clinic—diagnosis, management and complications of malignancy,” *Modern Techniques for Computer-Aided Melanoma Diagnosis*, 2011.
28. I. Maglogiannis, “Design and Implementation of a Calibrated Store and Forward Imaging System for Teledermatology,” *Journal of Medical Systems*, vol. 28, no. 5, pp. 455–467, 2004, https://doi.org/10.1023/B:JOMS.0000041172.70027.a0.
29. I. Maglogiannis and C. N. Doukas, “Overview of advanced computer vision systems for skin lesions characterization,” *IEEE transactions on information technology in biomedicine*, vol. 13, no. 5, pp. 721–733, 2009.
30. A. G. Manousaki *et al.*, “A simple digital image processing system to aid in melanoma diagnosis in an everyday melanocytic skin lesion unit. A preliminary report,” *International journal of dermatology*, vol. 45, no. 4, pp. 402–410, 2006.
31. S. E. Umbaugh, R. H. Moss, and W. V. Stoecker, “Applying artificial intelligence to the identification of variegated coloring in skin tumors,” *IEEE engineering in medicine and biology magazine*, vol. 10, no. 4, pp. 57–62, 1991.
32. M. Filho, Z. Ma, and J. M. R. S. Tavares, “A Review of the Quantification and Classification of Pigmented Skin Lesions: From Dedicated to Hand-Held Devices,” *J Med Syst*, vol. 39, no. 11, p. 177, 2015, https://doi.org/10.1007/s10916-015-0354-8.
33. S. Dreiseitl, L. Ohno-Machado, H. Kittler, S. Vinterbo, H. Billhardt, and M. Binder, “A comparison of machine learning methods for the diagnosis of pigmented skin lesions,” *Journal of biomedical informatics*, vol. 34, no. 1, pp. 28–36, 2001.
34. J. Sanders, B. Goldstein, D. Leotta, and K. Richards, “Image processing techniques for quantitative analysis of skin structures,” *Computer methods and programs in biomedicine*, vol. 59, no. 3, pp. 167–180, 1999.
35. S. Tomatis, A. Bono, C. Bartoli, G. Tragni, B. Farina, and R. Marchesini, “Image analysis in the RGB and HS colour planes for a computer-assisted diagnosis of cutaneous pigmented lesions, ” *Tumori*, vol. 84, no. 1, pp. 29–32, 1998.
36. A. Bono *et al.*, “The invisible colours of melanoma. A telespectrophotometric diagnostic approach on pigmented skin lesions,” *European Journal of Cancer*, vol. 32, no. 4, pp. 727–729, 1996.
37. B. Chwirot, S. Chwirot, J. Redziński, and Z. Michniewicz, “Detection of melanomas by digital imaging of spectrally resolved ultraviolet light-induced autofluorescence of human skin,” *European Journal of Cancer*, vol. 34, no. 11, pp. 1730–1734, 1998.
38. I. Maglogiannis and E. Zafiropoulos, “Utilizing support vector machines for the characterization of digital medical images,” *BMC Medical Informatics and Decision Making*, vol. 4, no. 4, 2004.
39. G. L. Hansen, E. M. Sparrow, J. Y. Kokate, K. J. Leland, and P. A. Iaizzo, “Wound status evaluation using color image processing,” *IEEE Transactions on Medical Imaging*, vol. 16, no. 1, pp. 78–86, 1997.
40. Z. Zhang, R. H. Moss, and W. V. Stoecker, “Neural networks skin tumor diagnostic system,” 2003, vol. 1, pp. 191–192.
41. K. Korotkov and R. Garcia, “Computerized analysis of pigmented skin lesions: a review,” *Artificial intelligence in medicine*, vol. 56, no. 2, pp. 69–90, 2012.
42. H. Motoyama, T. Tanaka, M. Tanaka, and H. Oka, “Feature of malignant melanoma based on color information,” 2004, vol. 1, pp. 230–233.
43. M. Herbin *et al.*, “Assessment of healing kinetics through true color image processing,” *IEEE Transactions on Medical Imaging*, vol. 12, no. 1, pp. 39–43, 1993.
44. W. Lohmann and E. Paul, “In situ detection of melanomas by fluorescence measurements,” *Naturwissenschaften*, vol. 75, no. 4, pp. 201–202, 1988.
45. J. C. Boldrick, C. J. Layton, J. Nguyen, and S. M. Swetter, “Evaluation of digital dermoscopy in a pigmented lesion clinic: clinician versus computer assessment of malignancy risk,” *Journal of the American Academy of Dermatology*, vol. 56, no. 3, pp. 417–421, 2007.
46. E. Lefevre, O. Colot, P. Vannoorenberghe, and D. de Brucq, “Knowledge modeling methods in the framework of evidence theory: an experimental comparison for melanoma detection,” 2000, vol. 4, pp. 2806–2811.
47. R. J. Stanley, R. H. Moss, W. Van Stoecker, and C. Aggarwal, “A fuzzy-based histogram analysis technique for skin lesion discrimination in dermatology clinical images,” *Computerized Medical Imaging and Graphics*, vol. 27, no. 5, pp. 387–396, 2003.
48. S. E. Umbaugh, Y.-S. Wei, and M. Zuke, “Feature extraction in image analysis. A program for facilitating data reduction in medical image classification,” *IEEE engineering in medicine and biology magazine*, vol. 16, no. 4, pp. 62–73, 1997.
49. M. Monisha, A. Suresh, and M. R. Rashmi, “Artificial Intelligence Based Skin Classification Using GMM,” *J Med Syst*, vol. 43, no. 1, p. 3, 2018, https://doi.org/10.1007/s10916-018-1112-5.

50. H. Ganster, P. Pinz, R. Rohrer, E. Wildling, M. Binder, and H. Kittler, “Automated melanoma recognition,” *IEEE transactions on medical imaging*, vol. 20, no. 3, pp. 233–239, 2001.
51. C. Grana, G. Pellacani, R. Cucchiara, and S. Seidenari, “A new algorithm for border description of polarized light surface microscopic images of pigmented skin lesions,” *IEEE Transactions on Medical Imaging*, vol. 22, no. 8, pp. 959–964, 2003.
52. P. Rubegni *et al.*, “Automated diagnosis of pigmented skin lesions, ” *International Journal of Cancer*, vol. 101, no. 6, pp. 576–580, 2002.
53. F. Ercal, A. Chawla, W. V. Stoecker, H.-C. Lee, and R. H. Moss, “Neural network diagnosis of malignant melanoma from color images,” *IEEE Transactions on biomedical engineering*, vol. 41, no. 9, pp. 837–845, 1994.
54. G. R. Lanckriet, T. De Bie, N. Cristianini, M. I. Jordan, and W. S. Noble, “A statistical framework for genomic data fusion,” *Bioinformatics*, vol. 20, no. 16, pp. 2626–2635, 2004.
55. J. Ye *et al.*, “Heterogeneous data fusion for alzheimer’s disease study,” 2008, pp. 1025–1033.
56. M. Kashani-Sabet *et al.*, “A multimarker prognostic assay for primary cutaneous melanoma,” *Clinical Cancer Research*, vol. 15, no. 22, pp. 6987–6992, 2009.
57. G. J. Mann *et al.*, “BRAF mutation, NRAS mutation, and the absence of an immune-related expressed gene profile predict poor outcome in patients with stage III melanoma,” *Journal of Investigative Dermatology*, vol. 133, no. 2, pp. 509–517, 2013.
58. B. E. G. Rothberg, M. B. Bracken, and D. L. Rimm, “Tissue biomarkers for prognosis in cutaneous melanoma: a systematic review and meta-analysis,” *Journal of the national cancer institute*, 2009.
59. Z. Xu, Y. Zhou, Y. Cao, T. L. Dinh, J. Wan, and M. Zhao, “Identification of candidate biomarkers and analysis of prognostic values in ovarian cancer by integrated bioinformatics analysis,” *Medical oncology (Northwood, London, England)*, vol. 33, no. 11, p. 130, 2016, https://doi.org/10.1007/s12032-016-0840-y.
60. G. T. Gibney, L. M. Weiner, and M. B. Atkins, “Predictive biomarkers for checkpoint inhibitor-based immunotherapy,” *The Lancet. Oncology*, vol. 17, no. 12, pp. e542–e551, 2016, https://doi.org/10.1016/s1470-2045(16)30406-5.
61. K. Moutselos, I. Maglogiannis, and A. Chatziioannou, “Integration of high-volume molecular and imaging data for composite biomarker discovery in the study of melanoma,” *BioMed research international*, vol. 2014, p. 145243, 2014, https://doi.org/10.1155/2014/145243.
62. I. Valavanis, I. Maglogiannis, and A. Chatziioannou, “Exploring robust diagnostic signatures for cutaneous melanoma utilizing genetic and imaging data,” *IEEE journal of biomedical and health informatics*, pp. 190–198, 2015.
63. *epsos*.
64. M. Maragoudakis and I. Maglogiannis, “Skin lesion diagnosis from images using novel ensemble classification techniques,” 2010, pp. 1–5.
65. I. Maglogiannis, S. Pavlopoulos, and D. Koutsouris, “An integrated computer supported acquisition, handling, and characterization system for pigmented skin lesions in dermatological images,” *IEEE Transactions on Information Technology in Biomedicine*, vol. 9, no. 1, pp. 86–98, 2005.
66. G. Kontogianni, O. Papadodima, I. Maglogiannis, K. Frangia-Tsivou, and A. Chatziioannou, “Integrative Bioinformatic Analysis of a Greek Epidemiological Cohort Provides Insight into the Pathogenesis of Primary Cutaneous Melanoma,” 2016.
67. G. Kontogianni, G. Piroti, I. Maglogiannis, A. Chatziioannou, and O. Papadodima, “Dissecting the Mutational Landscape of Cutaneous Melanoma: An Omic Analysis Based on Patients from Greece,” *Cancers*, vol. 10, no. 4, p. 96, 2018, https://doi.org/10.3390/cancers10040096.
68. E. Cerami *et al.*, “The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data,” 2012.
69. J. Gao *et al.*, “Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal,” *Science signaling*, vol. 6, no. 269, p. pl1, 2013, https://doi.org/10.1126/scisignal.2004088.
70. R. D. Melamed *et al.*, “Genomic characterization of dysplastic nevi unveils implications for diagnosis of melanoma,” *Journal of Investigative Dermatology*, vol. 137, no. 4, pp. 905–909, 2017.
71. I. A. Adzhubei *et al.*, “A method and server for predicting damaging missense mutations,” in *Nat Methods*, vol. 7, United States, 2010, pp. 248–9.
72. T. Koutsandreas, I. Binenbaum, E. Pilalis, I. Valavanis, O. Papadodima, and A. Chatziioannou, “Analyzing and visualizing genomic complexity for the derivation of the emergent molecular networks,” *International Journal of Monitoring and Surveillance Technologies Research (IJMSTR)*, vol. 4, no. 2, pp. 30–49, 2016.
73. M. Ashburner *et al.*, “Gene ontology: tool for the unification of biology. The Gene Ontology Consortium,” *Nature genetics*, vol. 25, no. 1, pp. 25–9, 2000, https://doi.org/10.1038/75556.
74. The Gene Ontology Consortium, “The Gene Ontology Resource: 20 years and still GOing strong,” *Nucleic Acids Res.*, vol. 47, no. D1, pp. D330–D338, 2019, https://doi.org/10.1093/nar/gky1055.
75. A. Fabregat *et al.*, “The Reactome Pathway Knowledgebase,” *Nucleic Acids Res.*, vol. 46, no. D1, pp. D649–D655, 2018, https://doi.org/10.1093/nar/gkx1132.
76. N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, “SMOTE: synthetic minority over-sampling technique,” *Journal of artificial intelligence research*, vol. 16, pp. 321–357, 2002.
77. X. Chen and H. Ishwaran, “Random forests for genomic data analysis,” *Genomics*, vol. 99, no. 6, pp. 323–9, 2012, https://doi.org/10.1016/j.ygeno.2012.04.003.
78. R. Development (2011) “Core TeamR: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing,” ISBN 3-900051-07-0. Available: h ttp://www. R-project. org.
79. M. Kuhn, “Caret: classification and regression training,” *Astrophysics Source Code Library*, 2015.
80. L. Torgo, *Data mining with R: learning with case studies*. CRC press, 2016.
81. X. Robin *et al.*, “pROC: an open-source package for R and S+ to analyze and compare ROC curves,” *BMC bioinformatics*, vol. 12, no. 1, p. 77, 2011.
82. K. Hajian-Tilaki, “Receiver Operating Characteristic (ROC) Curve Analysis for Medical Diagnostic Test Evaluation,” *Caspian journal of internal medicine*, vol. 4, no. 2, pp. 627–35, Spring 2013.
83. Brooke, J., “SUS – A Quick and Dirty Usability Scale,” in *Usability Evaluation in Industry*, vol. 194, 1996, pp. 4–7.
84. Brooke, J., “SUS: a retrospective,” *Journal of usability studies*, vol. 8, no. 2, pp. 29–40, 2013.

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.