A Complementary Sensing Platform for a holistic approach to Allergic Rhinitis monitoring

Andreas Bardoutsos[1,2], Giorgos Matzarapis[1,2], Sotiris Nikoletseas[1,2],
Paul G. Spirakis[1,2,3], and Pantelis Tzamalis[1,2]

[1]Computer Engineering and Informatics Department, University of Patras, Patras, Greece
[2]Computer Technology Institute and Press "Diophantus" (CTI), Patras, Greece
[3]Department of Computer Science, University of Liverpool, UK
bardoutsos@ceid.upatras.gr, matzarapis@ceid.upatras.gr, nikole@cti.gr
p.spirakis@liverpool.ac.uk, , tzamalis@ceid.upatras.gr

***Abstract*—Allergic diseases and, in particular, allergic rhinitis are among the most common chronic diseases, inducing disturbances in daily activities. They are caused primarily by the pollens of allergenic plants and symptoms can deteriorate due to various ambient conditions which work as irritants, such as humidity. In this paper, we present the development of an eHealth/mHealth holistic platform that utilizes the technologies of Internet of Things (IoT), Mobile Crowdsensing (MCS), Social Networking Services, Natural Language Processing (NLP), and Machine Learning (ML), in order to work as a sentinel and disease prevention tool for patients with allergic rhinitis symptoms. By efficiently combining human with machine intelligence, we provide a complementary sensing method for the comprehensive and large-scale monitoring of the disease in broad regions, and in real-time. Moreover, the users of our platform are encouraged to engage in the sensing process through a personalized health monitoring system in order to keep a constant awareness of their symptoms and, thus, deliver a successful adherence to their treatment. As an important use case, we adapted our platform to the USA region, but it can be easily extended to any other area with minor modifications. The design and complete implementation of our platform has been performed and validated in close cooperation with well-recognized academic medical doctors based in Greece who specialize in the control of allergic diseases (and rhinitis in particular) and provided valuable insights and detailed requirements analysis about the functionality and usability of the platform. To the best of our knowledge, this is the first study that examines allergic rhinitis monitoring in a complementary manner and on large scale, with the utilization of hybrid data sources.**

***Index Terms*—allergic rhinitis, ehealth, mhealth, IoT, sensors, social media, mobile crowdsensing, machine learning, natural language processing**

I. Introduction

Millions of people suffer from environmental allergies. For instance, focusing on the United States of America, 1 out of 5 Americans are diagnosed as allergic, which has a serious impact on individuals' Quality of Life (QOL) and economy, as for example, resulting almost in 4 million missed workdays per year [1]! Thus, surveillance systems for monitoring and controlling such disease outbreaks are in high-demand.

The most common type of allergy caused by environmental factors is allergic rhinitis which is mostly caused by pollen [2]. Pollen grains are small, light, and dry which keeps them airborne even over long distances. Their ingestion, however, triggers immune system reactions. Furthermore, it must be noted that pollen concentrations in the air are highly correlated with various environmental factors, and the novel interactions with them results also in allergy prevalence [3]. Generally, ambient conditions such as humidity or dust are feasibly monitored through ubiquitous sensor networks which are indeed a valuable tool for the monitoring of diverse phenomena of the physical world. On the other hand, sensors related to allergens monitoring are very expensive to be manufactured, placed, and maintained, thus, the spatial coverage of such systems is quite restricted. Therefore, the monitoring of the pollen and as a consequence, the allergy and its symptoms onsets, is tricky.

As a response to the ongoing situation, crowd-intelligence technologies can complement the functionality of sensor networks whenever the last is not always available in some areas or too costly to deploy. First, the utilization of Mobile Health Crowdsensing (MHCS) through the Healthcare 4.0 topic is indicative for supporting such health care practices [4]. Specifically, the exploitation of smart mobile devices' sensors, such as smartphones, combined with human intelligence, provides the ability to the government, organizations, authorities, and relevant stakeholders to develop strategic tools for efficient monitoring, studying, and decision-making towards the confrontation of diverse public health challenges. Updates and warnings of a possible disease outbreak can be also adopted, the patients can be stratified based on their disease/treatment, and the allergy exacerbation can be early prevented through the development of such infrastructure, in a timely manner. Thus, it is obvious how important such systems are for maintaining the balance of public and personal health. Secondly, an additional pool of data can also be obtained from the vast popularity of Social Networking Services. Social media engagement reflects a general interest in a range of issues, including health conditions and concerns. Hence, indicators regarding a disease activity could be derived by studying the interactions through these networks [5] [6]. As a result, these methods exploit the ubiquity of the crowd to gain insights, in real-time when objective sensing is not available.

Definitely, though, the combination of objective information from the sensor networks with the participatory inputs from the crowd, and the exploitation of the information shared through Social Media, turns large-scale monitoring into a realization.

Our contribution. In this paper, we negotiate the development of a hybrid-inputs eHealth and mHealth platform for the large-scale monitoring and outbreaks detection of allergic rhinitis, pollen allergens, and irritants. As an important special case, we focused on the USA territory. Our platform comprises various components that work complementary to each other to provide a real-time comprehensive overview of the disease's impact on population health, as well as the surveillance of the pollen onsets. Diverse input data from users' participation, Social Networking Services, and sensor measurements are efficiently combined and analyzed to provide a universal output of the ongoing state of disease outbreak, even for places where any of the inputs are lack presence. Moreover, the patient's stratification and adherence to treatment are promoted through a personalized health monitoring system. The user is able to monitor her health status through time and trace her past locations where allergic symptoms had aggravated. Additionally, it is worth mentioning that every step of the implementation procedure has been performed under the supervision of experienced physicians that verified its functionality and usability in the context of a large-scale funded research-project, titled "Personal Allergy Tracer" [7]. Finally, our platform can easily be modified and used for other areas outside the US, with minor configurations, and, over its interface, the users are able to access its features from any smart device (e.g. laptops, smartphones, tablets). Such configurations are the identified allergens that affect the citizens over a region based on its geographical position, as well as the processing and analysis of the corresponding language that is mostly used in this particular area of interest.

II. Related Work

Remarkable research approaches exploit the technologies of sensor networks, MCS, and Social Networking Services lately as an efficient method to study the confrontation of diverse health-related problems, the monitoring of diseases' outbreaks, and finally, improve decision-making policies, including allergies too. For instance, Pollen Web Application[1], is an online platform that provides information and warns allergic patients regarding the spatial pollen status in the United States. The allergens' exacerbation is based on sensor data measurements that originate from a network of *Rotorod sensors* [8] that are distributed across the country. From the MCS point of view, [9] and [10] are oriented towards monitoring atmospheric and noise pollution correspondingly. The recording of the locations where the sensing is submitted using geolocation data is a key aspect of both systems. Others, like [11], utilized Twitter[2] content and Machine Learning (ML) to monitor allergies in a spatiotemporal and proved that Social Networks activity related to allergies is highly correlated with real metrics of the pollen volume.

From our research approaches perspective, in [3], a hybrid eHealth/mHealth MCS system was developed for the spatiotemporal monitoring of allergens outbreaks, the symptoms exacerbations, and the relevant irritants that affect them regarding the Greek territory. A mechanism was also deployed for the patient's health status surveillance. Specifically, in a poster paper in [12], we presented a preliminary version of our *envisioned* platform. It was our first partial attempt for the spatiotemporal monitoring of allergic rhinitis symptoms exacerbations by exploiting the hybrid inputs of the power of the crowd and sensor networks. However, only the utilization of Social Media posts took place in this study with the deployment of text-mining techniques, where we used some ideas from our previous work pertaining to cybercrime activity surveillance as presented in detail in [13].

Despite the contribution of the above studies in the control of allergic diseases, we noticed that none of the analyses derived from them could achieve efficient spatiotemporal coverage monitoring on a large scale, due to the inability to handle the inherently dynamic nature (or even complete lack) of data sources from various regions effectively. Thus, we went after combining the mining of a set of *"senses"* through a complementariness of analysis from various sources.

Novelty. In our work here, our entire and *novel* research approach is extended by efficiently implementing our envisioned allergic rhinitis surveillance and sentinel platform, for large region-scale. Despite the challenges, the various components bind perfectly together in the procedure of collection, processing, and analysis of the hybrid inputs from *sensor data*, *Twitter posts*, and *participatory sensing* based on patients' attitudes related to their allergic symptoms combined with geolocation information. Thus, the disease outbreaks are entirely monitored as even for locations where such of the inputs is not available, the allergic activity is inferred through the rest of the inputs. Finally, the platform is deployed and validated with a focus on the USA area, since it provides a wide range of data availability, both in sensor measurements and citizens' contribution to the Twitter platform as its use is very popular in the US unlike other countries, such as Greece. However, with minor modifications, the platform is able to provide valuable information for any spatial region. As far as we know, this is the first time an eHealth/mHealth monitoring application utilizes the complementariness of sensing analysis to provide valuable information for disease exacerbation monitoring and control, on a large scale.

Roadmap. The rest of the paper is organized as follows. Section III introduces the proposed platform architecture and highlights its components. In section IV, the diverse inputs to our platform are analyzed. The main analysis processes and outcomes are described in Section V. Additional features of the platform are overviewed in section VI. Our future work is described in the final section where the paper is concluded.

[1]https://www.pollen.com/

[2]https://twitter.com/

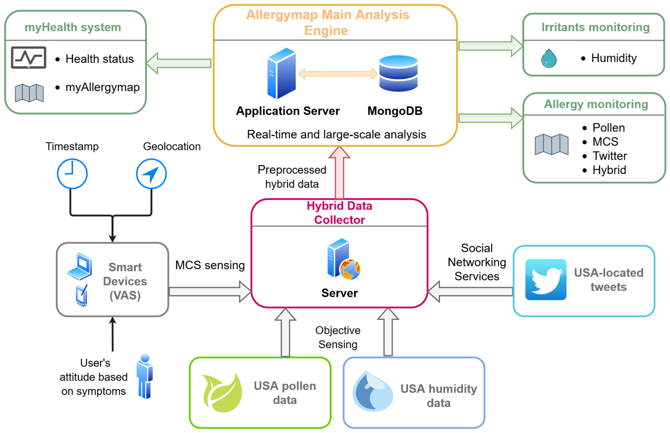

Fig. 1. Platform Architecture.

III. Platform Architecture

Our overall approach is depicted in Figure 1 where each component that comprises our hybrid platform is presented. Our platform gathers and stores information from heterogeneous sources, with the use of the Hybrid Data Collector component. The main input entities are highlighted below:

- *Subjective*: It is based on the human participation where the users' personal assessment of allergic activity combined with spatial and temporal identifiers is recorded through a Web interface, accessible by any smart device.
- *Objective*: The sensor data is collected regarding pollen concentration levels and humidity levels regarding the USA territory.
- *Social Media*: Through this component, text posts from Twitter which are related to allergic rhinitis and pollen onsets are collected.

Multiple analysis is exported by this data, as all the aforementioned inputs to the platform are combined in the Main Analysis Engine. The outcomes refer not only to the patients' participation but also to any stakeholder that is interested in disease monitoring. More specifically, the platform outputs are:

- *Allergic rhinitis and humidity monitoring*: The monitoring of the ongoing disease effect is achieved through spatiotemporal representation of each of the objective inputs, the subjective inputs, and their hybrid analysis. In particular, activity related to pollen allergens levels and allergic rhinitis symptoms intensity is visualized through interactive maps, as well as time-series and categorical analysis plots. The monitoring of the humidity irritant levels in that US states' regions is calculated from sensor data and is also depicted in the relevant map.
- *The myHealth system*: It defines a personalized output that provides information to each user that participates in the sensing procedure about the course of her health and the locations her symptoms occurred, as well as access to a personal allergy profile.

The development of our platform is based upon the Python Flask microframework[3]. The Web Server Gateway Interface (WSGI) forwards the requests of our Web server to the application. The User Interface (UI) is designed by utilizing the Bootstrap templates[4]. Moreover, the non-relational database MongoDB[5] was selected as a data warehouse due to its unique specifications regarding timely and spatially annotated data manipulation. Last but not least, it is worth mentioning that preserving the users' privacy is an unconditional premise for our platform development. Thus, the users are informed of the data that is recorded regarding their activity, all the analysis procedures process the data anonymously, and security functions have been deployed as presented in *Appendix A*.

IV. The Data Collection Process

The data collection process in our platform is distributed across the various input components. Each of them corresponds to a particular aspect of disease monitoring. The functionality of each of those components is described below.

A. User Profiling & participatory sensing (MCS)

In order to ensure that data originates mainly from allergic patients and further ensure data reliability, registration is mandatory for participation in the data collection process. The registration consists of three stages.

At the first stage, the user's basic information and authentication data are required for her account creation. The second stage comprises a series of questionnaires that have to be fulfilled for the creation of a personalized allergic profile. More specifically, in the first section, she is prompted to provide information about 1) the gender, 2) her family medical history related to allergies, 3) the affections of allergic symptoms in her daily life, 4) if receiving immunotherapy, to specify the duration, and 5) the duration of receiving allergy-specific medication. The next step refers to recognizing and submit her most common allergic symptoms and specify the allergens that affect her, as long as they are previously known, while the last questionnaire, is oriented towards specific information about the medication she receives. Furthermore, it must be noted that after the initial steps of registration the user has always access to both her personal and allergic profiling information which can be changed at any time, as Fig. 10 of Appendix B depicts.

The last stage is where human intelligence is embodied in the sensing procedure. The submission of daily symptoms through the VAS tool [14] takes place as shown in Fig. 2. It is, actually, the user involvement in the data collection process of that component. The records correspond to a psychometric response scale relevant to the user's attitude regarding the perceived intensity of her symptoms at a particular moment. The scale ranges from zero to twelve (0-12), where zero corresponds to the absence of symptoms and twelve to a severe allergic outbreak and thus high intolerance. Additionally, each VAS recording is also composed of time and geolocation

[3]https://github.com/pallets/flask
[4]https://getbootstrap.com/
[5]https://www.mongodb.com/

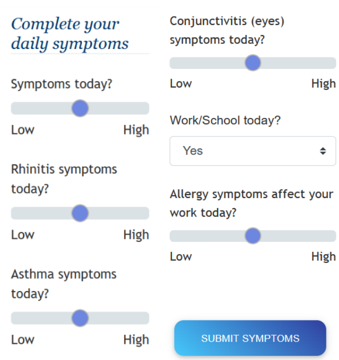

Fig. 2. The VAS tool.

data with the latter one obtained by exploiting the HTML5 Geolocation API[6]. Nevertheless, in the case of a not supported or disabled positioning determination, a map interface is provided to the user (Fig. 11, *Appendix B*) for the purpose of the manual location annotation. Moreover, the tool usage can not only be accessible during the registration procedure but whenever a user deals with symptoms exacerbations.

Finally, by exploiting the in-browser cookies functionality, the user is notified to update her health status on a daily basis and is also reminded to inspect her allergic rhinitis profiling information in regular time periods in case a modification is desirable or requested As a result, the user does not neglect to interact with the platform and her devotion and compliance to the treatment are significantly improved.

B. Objective Sensing

The platform's objective sensing in various regions comprises two inputs, the pollen measurements, and the humidity levels which compose the irritant onsets' knowledge of our study. First, the pollen objective inputs originate from two main sources. Pollen data is retrieved from the AQVIA Open-Source API[7] that distributes measurements obtained by the Rotorod Sampler sensors. These measurements indicate how much pollen is in the air and refer to the concentration of either all allergens' pollen or the specific ones originating from a particular allergen in the air at a particular time in a given region. It is expressed in pollen grains per cubic meter for a period of 24 hours and its value ranges between 0 and 12. Secondly, humidity consists a known irritant, since humidity levels directly affect the severity of allergic symptoms as well as individuals' wellness [15]. Hence, the monitoring of humidity level in the air is highly important and, for that reason, it is specified as an additional objective input to the platform. The corresponding measurements are obtained from the Weather API[8] provided by OpenWeatherMap[9].

C. Indirect crowdsourcing sensing

Twitter is our primary source of data in this component, as it provides access to massive up-to-date and real-time information generated continuously by users. Allergic rhinitis activity monitoring is performed on a global-scale by collecting tweets that include a set of keywords and hashtags related to our topic, as referenced in Table II of *Appendix B*. A tweet is accepted only if at least one of the keywords provided in this Table is detected. The data is collected in real-time through the utilization of the Streaming OAuth 1.0a API[10] with a maximum response time of 30 seconds from the time the tweet is published. Finally, the Python *tweepy*[11] library is used to communicate with Twitter's Streaming API[12].

D. Hybrid Data Collector

Our platform gathers direct information from the users through the VAS tool, indirect crowdsourced inputs from allergic-rhinitis-related tweets, and objective sensing measurements from pollen and humidity levels. All this data originates from different sources and each one needs different manipulation for the communication and reception of the corresponding data. For that purpose, the Hybrid Data Collector tool has been implemented. Its role is to establish the connection with the corresponding data warehouse or API (AQVIA, OpenWeather, and Twitter) in order to collect the relevant information and prepare the data for insertion to the Main Analysis Engine where the knowledge extraction takes place. It is important to mention that our platform can be expanded to support other input sources, simply through the addition of the connection configuration with an additional data source.

V. Main Analysis Engine and Outcomes

The main processing unit of our platform is the *Main Analysis Engine*. Extended statistical analysis is conducted or ML models are applied to the data that is gathered from the various input components through the *Hybrid Data Collector*, and each component has the ability to complement the other in a hybrid sensing manner. Thus, it establishes a holistic approach for allergic rhinitis monitoring and its relevant irritants exacerbations both on a large- and individual-scale, in real-time. Interactive visualization techniques, such as the Google Maps Platform[13], as well as the Leaflet[14] and the Highcharts[15] JavaScript libraries, are utilized also in order to provide a user-friendly environment where cognitive knowledge and better decision-making can be extracted from the processed results. In this section, each discrete procedure that takes place during the analysis is described and the corresponding outcomes are presented.

A. Pollen and Humidity monitoring

Pollen counts that are stored in our database originate from 968 sensor stations all over the USA, and refer to *current*, *historic*, and *forecasted* pollen allergens levels. As depicted

[6]https://www.w3.org/TR/geolocation-API/
[7]https://github.com/bachya/pyiqvia/
[8]https://openweathermap.org/api
[9]https://openweathermap.org/
[10]https://developer.twitter.com/en/docs/authentication/oauth-1-0a
[11]https://www.tweepy.org/
[12]https://developer.twitter.com/en/docs
[13]https://developers.google.com/maps
[14]https://leafletjs.com/
[15]https://www.highcharts.com/

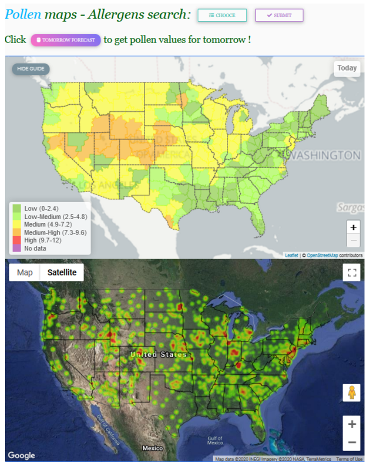

Fig. 3. Pollen onset monitoring from sensor measurements.

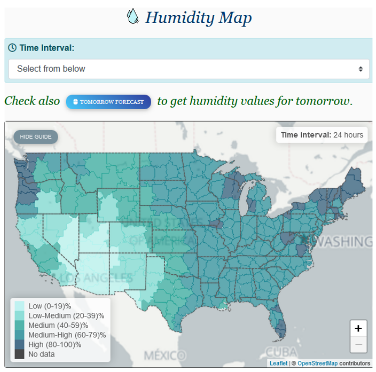

Fig. 4. Humidity monitoring based on sensor measurements.

in Fig. 3, there are two different outcomes that our analysis of these objective measurements yields. In the first case, the obtained data is presented in a regional manner, through an interactive choropleth map. Following the spatial segmentation of the USA area, such in *Pollen.com* platform, 245 areas are specified. Each discrete region's borders occur from a collection of coordinates (longitude, latitude) that altogether shape a polygon and delimit the corresponding geographical territory. Subsequently, pollen densities are calculated as the average of all the sensors' measurements that are located within each of those areas. It should be mentioned though, that diverse allergens contribute to the overall pollen levels. However, the platform's User Interface (UI) provides through its corresponding menu the ability to the user to choose between a set of pollen allergens which can be plotted dynamically to the map. The second way regarding pollen detection representation is a heatmap that provides higher accuracy as it occurs according to the zip code of the area that a sensor is located. Color intensity corresponds to the level of pollen levels and the density of located sensors is presented from each mark diameter. As someone can infer from this plot, indeed, the spatial coverage of the US area is not holistically feasible. Thus, we betake to the complementary sensing implementation for the allergic rhinitis monitoring by utilizing the power of crowdsourced-intelligence as presented in the following section.

Finally, following the same principles as in pollen denseness spatial representation development, humidity is also presented in the corresponding map as our allergic-rhinitis-related irritant's monitoring procedure, which is visualized in Fig. 4. A representative humidity value is calculated as the mean of the sensors measurements located inside each of the USA areas' spatial polygons as they have been previously defined.

B. Crowd-intelligence

MCS and Twitter users represent the sentinels for the onsets of the allergic rhinitis season. Hence, the combination of these sources highlights perfectly the intensity of allergic symptoms, exclusively from the human factor. In order to combine this information, an extensive analysis for each human-intelligence data source is necessary to be applied.

Participatory (MCS) sensing analysis. The users' VAS tool submissions regarding their attitude based on allergic symptoms are aggregated in combination with temporal and geolocation data to provide a human-centered approach to the allergic activity. Following the methodology applied in pollen allergens monitoring, the intensity of the allergic symptoms in a region is calculated as the average of VAS counts intensity combined with the density of the users' inputs in a specific spatial region and time interval. Moreover, another key functionality of this map resides in the capability to visualize the allergic symptoms activity caused specifically by particular allergens from which each user is affected. Specifically, each user is sensitive to specific allergens. Thus, the perceived intensity of the user's symptoms infers the presence of these particular allergens that are the cause of her irritation. It should be noted though, that the users' privacy is our concern. Hence, the analyzed traces are completely anonymized and cannot infer a user location or leak personal information as the location accuracy is presented inside a 400-meter radius of the initial user position.

Social Networking services analysis. The indirect human participation of our platform derives from the collected Social Networks' activity. Until now, 135,703,011 tweets related to allergic rhinitis have been collected to the database. These posts come from a worldwide range and are not all explicitly relevant to the actual onset of allergic rhinitis and its symptoms. For this purpose, various methods, algorithms, ML models, and text-mining techniques (NLP or statistical) are applied in the various steps of the raw text cleansing, processing, and analysis to achieve better mining from this source of information which will be described below.

Due to the fact that geographic coordinates are available on Twitter only for a limited number of users (i.e. $< 2\%$) because of their privacy concerns, the localization of tweets is a challenging task. A two-step approach is therefore followed for the identification of their location in the US as proposed in our previous work [12]. Firstly, the wide-area tweets are derived according to the timezone they were posted. Secondly, the location is either retrieved from the sensor-based coordinates

TABLE I
FEATURES PRODUCED FROM ADVANCED TEXT MINING.

Category	Features
Morphological	Length in characters, tokens, included URLs, hashtags, certain punctuation marks ('?', '!', '?!') and total punctuation marks, emojis, periods, stopwords, words, vowels, consonants, upper and lower case characters, digits, letters, average number of characters per word, average number of words per period, length of the longest sequence of vowels and consonants in a word.
Part of Speech (PoS)	Corresponding occurrences of verbs, entities, pronouns, determiners, and adverbs in the tweet text
Semantic	Positive, neutral and negative sentiment derived from text and emojis
Other	TTR, tweet entropy, TF-IDF

if they are available in the tweet's object or from the text-based profile location that is publicly visible on her page. In the latter one, complex regular expressions have applied that match the state's code with the raw text e.g. New York, NY, etc. which is then assigned to the corresponding geocoding information for analysis and plotting the data in a spatial manner. Moreover, the user profiles that shared the collected content are checked if they correspond to real users by utilizing the Botometer Python API[16] for the bot detection.

Another analysis step is the identification of the relevant to our interest collected tweets by developing a bag-of-words supervised ML approach. For that purpose, a tweets annotation tool was developed to create a dataset of "relevant" and "irrelevant" labeled to our topic posts. Then, a set of 37 features is extracted from the raw text which is related to morphological level, Type to Token Ratio (TTR), the tweet entropy, a number of Part of Speech (PoS) features, a wide array of semantic features based on the sentiment derived from the text and emojis. Moreover, TF-IDF (term frequency-inverse document frequency) statistical measure is applied for the generation of another set of features for advanced information retrieval [16]. All the extracted features are shown in Table I. Here, it is worth pointing out that for the generation of some features, processes of tokens' *stemming* and *lemmatization* are applied. The features are then standardized by removing the mean and scaling to unit variance, and finally, for performance issues, the dataset is projected to a lower-dimensional space by using Principal Component Analysis (PCA) to speed up the machine learning algorithm. Subsequently, various supervised ML models are trained, with the Support Vector Machines (SVM) providing the best results with an accuracy of 92%. Accordingly, the context of the newly retrieved tweets is automatically classified if it corresponds to our focus of concern. Finally, for the above-described procedure, the following Python libraries and

[16]https://github.com/IUNetSci/botometer-python

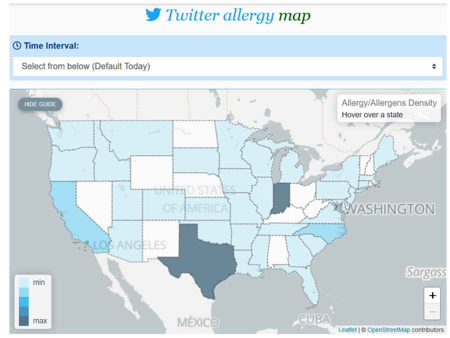

Fig. 5. Allergic rhinitis spatiotemporal mining based on Twitter posts.

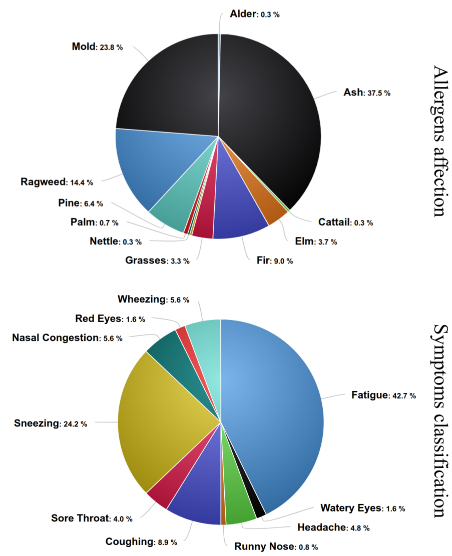

Fig. 6. Distribution of mentioned allergic rhinitis symptoms and allergens on the collected tweets.

APIs are utilized: a) *re*[17], *NLTK*[18], and *spaCy*[19] for the NLP processes, and, b) *scikit-learn*[20] for the scaling, dimensionality reduction, and training tasks.

The outcome of the aforementioned analysis is the spatiotemporal mining of the allergic rhinitis levels monitoring in the US. The map of Fig. 5 depicts the extracted knowledge which is presented to each state accordingly and anonymized as privacy awareness is major to our platform development. Moreover, except for the daily activity visualization, the visitor of our platform is also able to perform a search for a specific time interval through the corresponding menu. In that case, the average number of daily tweets from the selected period is dynamically calculated and depicted. Additionally, better insights are gained through a categorical text analysis that yields the distribution of symptoms, as well as the allergens' affection to the US citizens over the last three months as shown in Fig. 6. Based on these plots, one can infer that fatigue,

[17]https://docs.python.org/3/library/re.html
[18]http://www.nltk.org/
[19]https://spacy.io/
[20]https://scikit-learn.org/stable/

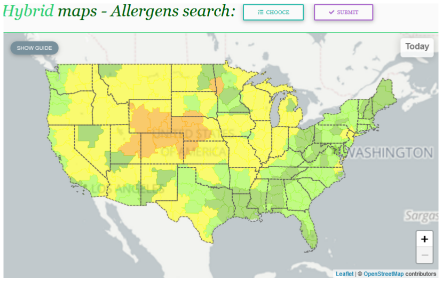

Fig. 7. The all-components hybrid spatiotemporal mining.

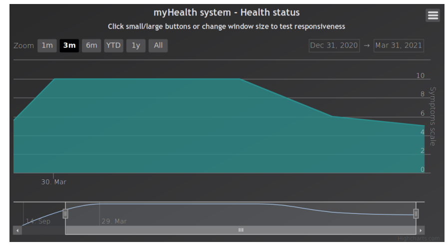

Fig. 8. myHealth system: The user's health status based on her historical attitude that occurred from her symptoms.

sneezing, and cough are the most common symptoms, while ash, mold, and ragweed are the ones that affect patients most. Finally, the time-series graph of Fig. 12 in *Appendix B* presents the number of allergic-rhinitis relevant tweets count over the course of time. Thus, these plots can alert and inform citizens, authorities, and stakeholders about the disease exacerbation and help the procedure of the decision making.

Hybrid Crowdsourcing. This kind of spatiotemporal analysis requires the calculation of a representative value, which raises two main challenges. The users' submissions from the VAS tool are on a constant scale between 0 and 12. On the other hand, the allergic activity in Twitter is calculated in accordance with tweet counts. Thus, there is no constant upper bound. As a result, the tweet counts should be transformed to a comparable size to the MCS inputs' scales. For that reason, a maximum value is specified as the maximum number of tweets detected within a day, in a two-week time period, for each state. Thus, considering this value as the Twitter scale maximum, normalization to the other inputs scale is calculated through the min-max normalization (Equation 1).

$$T = \frac{tweets\_today - min\_tweets}{max\_tweets - min\_tweets} * (ref\_max - ref\_min) + ref\_min \tag{1}$$

More precisely, T is the normalized value of allergy intensity according to the number of related tweets. $tweets\_today$ is defined as the daily number of tweets defined per state. max_tweets and min_tweets correspond to the maximum and minimum count of tweets per day over a two weeks time period. ref_max and ref_min are the upper and lower bounds of the reference data (0-12 in our case). The second challenge derives from the fact that the tweets are assigned to a state area in contrast with the rest of the sources that are assigned to 245 discrete regions. Thus, each state is matched to the areas it encloses and the hybrid score is calculated as a weighted average of the aforementioned inputs for each region. The result of this procedure is shown in Fig. 9 of *Appendix B*.

C. Holistic monitoring with Complementary Sensing

Finally, Fig. 7 depicts the outcome of the combined, hybrid analysis of the post-processed allergic-rhinitis information retrieved from all the objective and subjective processing procedures (MCS, pollen sensors counts, and Twitter). For the calculation of the hybrid effect of those sources, the normalization procedure as presented above is followed. It is important to mention that the contribution of each input to a final value that represents the allergic activities' intensity in a particular area is defined through a weight vector. This vector is calculated according to the availability and reliability of each source. Objective inputs have a higher contribution as they are considered the most reliable input. However, even in the absence of pollen sensors, the allergic-rhinitis score is derived from the rest of the inputs. The same applies to the absence of any of the other inputs, thereby accomplishing our complementary sensing paradigm.

Beyond that, the principal concept is to fill the gap between the human sense and the real environmental conditions. Considering the human factor, the intensity of allergic symptoms is completely different for each individual. Moreover, local weather conditions dramatically affect the perception of the symptoms. On the other hand, the sensors provide indisputable measurements of environmental conditions. Hence, human perspective and sensing technology complement each other to provide a holistic perception of the on-going disease impact.

VI. Motivation Features

Another benefit of the MCS in healthcare is the ability to increase the monitoring of the health progress of the citizens. Hence, the public should be motivated to participate in such initiatives through various incentives. For that purpose, the *myHealth* system is implemented and acts as an incentivization mechanism that aims to improve the patient's adherence to treatment by providing features that are based on the submitted recordings of the VAS tool. First, the user has access to a time-series graph (Fig. 8) that depicts her health status based on how the symptoms affect her attitude through time. Navigation to previous records and extraction of the plot for her own purposes is also supported. The second feature is the *myAllergyMap*, where each patient has access to her allergic traces heatmap (Fig. 13, *Appendix B*). Thus, a complete personalized history of locations that are associated with symptoms exacerbation is recorded and yields a useful tool that assists users in the avoidance of such places. Thirdly, a personal allergic profile is maintained for each user, as it occurred during the registration process. The submitted data defines an important reference for the efficacy of the user treatment. All the above information can also be extracted and shared with the relevant physician who is responsible for the patient's treatment monitoring. Moreover, as mention in section IV-A, the users are encouraged to update their information every two to four weeks, in case a change occurs.

Additionally, creating a user-friendly interface attracts users' interest and keeps them motivated to participate in the sensing procedure. Consequently, a section of the Web Application contains several informational allergy-related topics and articles that are provided via RSS feed[21]. Finally, a discreet articles column is displayed with dynamically renewed content, regarding allergy, treatment, and other on-going health affairs.

VII. Conclusion and Future Work

In the context of this work, an online platform was developed for the holistic spatiotemporal recording, detection, and comprehensive monitoring of allergic-rhinitis exacerbation, as well as relevant irritants that aggravate the disease symptoms, in real-time. The knowledge is provided with the efficient binding and cooperation of the individual components and the corresponding technologies they comprise. These are responsible for the collection, processing, analysis, and visualization of hybrid objective and subjective data sources, and, finally, the relevant information is extracted to interactive graph and map representations. Our platform has been fully implemented, deployed, tested, and validated in the United States of America and is concluded that is fully functional and can provide, in real-time, valuable information related to the disease onsets, as our collaboration with medical experts has verified the platform's usability in this real-world testing scenario. Thus, the complementary sensing through the combination of human and machine intelligence is now achieved and the platform can deliver to the interested stakeholders an even more accurate view of disease conditions in the area of their interest. Furthermore, the users can keep a record of their personalized allergic activity for the purpose of sharing them with their physician, if required.

Our future work focuses on collecting data from even more sources and update the methods of intelligent analysis, such as using Deep Learning methods, for the purpose of deriving further meaningful context. Regarding the Twitter data collection, adding the localization-based information at a city level instead of only the state levels is desired. In addition, it is our purpose to take into consideration new technologies and infrastructure architectures that would enable easier maintenance and impute scalability to our platform, like the containerization of the components. This will lead to an increased level of analysis towards the study and elimination of the effects of allergic rhinitis in order to provide improvements to the quality of patients' life.

Acknowledgment

This research has been co-financed by the European Union and Greek national funds through the Operational Program Competitiveness, Entrepreneurship and Innovation, under the call RESEARCH – CREATE – INNOVATE (project code: T1EDK 02436, project name: Personal Allergy Tracer).

[21] https://www.rssboard.org/

References

[1] "Allergy statistics in the us," Jan 2021. [Online]. Available: https://allergyasthmanetwork.org/allergies/allergy-statistics/

[2] R. A. Nathan, E. O. Meltzer, J. C. Seiner, and W. Storms, "Prevalence of allergic rhinitis in the united states," *Journal of allergy and clinical immunology*, vol. 99, no. 6, pp. S808–S814, 1997.

[3] L. A. Kalogiros, K. Lagouvardos, S. Nikoletseas, N. Papadopoulos, and P. Tzamalis, "Allergymap: a hybrid mhealth mobile crowdsensing system for allergic diseases epidemiology: a multidisciplinary case study," in *2018 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*. IEEE, 2018, pp. 597–602.

[4] T. O. Tokosi and M. Twum-Darko, "Mobile health crowdsensing (mhcs) intervention on chronic disease awareness: Protocol for a systematic review," *JMIR Research Protocols*, vol. 10, no. 3, p. e24589, 2021.

[5] A. Signorini, A. M. Segre, and P. M. Polgreen, "The use of twitter to track levels of disease activity and public concern in the u.s. during the influenza a h1n1 pandemic," *PLOS ONE*, vol. 6, no. 5, pp. 1–10, 05 2011. [Online]. Available: https://doi.org/10.1371/journal.pone.0019467

[6] K. Lee, A. Agrawal, and A. Choudhary, "Real-time disease surveillance using twitter data: Demonstration on flu and cancer," in *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '13. New York, NY, USA: Association for Computing Machinery, 2013, pp. 1474–1477. [Online]. Available: https://doi.org/10.1145/2487575.2487709

[7] "Personal Allergy Tracer," http://pat.vidavo.eu/, 2021, [Online; accessed 30-March-2021].

[8] D. A. Frenz and B. L. Guthrie, "A rapid, reproducible method for coating rotorod sampler collector rods with silicone grease," *Annals of Allergy, Asthma & Immunology*, vol. 87, no. 5, pp. 390 – 393, 2001. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1081120610629201

[9] M. Mun, S. Reddy, K. Shilton, N. Yau, J. Burke, D. Estrin, M. Hansen, E. Howard, R. West, and P. Boda, "Peir, the personal environmental impact report, as a platform for participatory sensing systems research," in *Proceedings of the 7th International Conference on Mobile Systems, Applications, and Services*, ser. MobiSys '09. New York, NY, USA: Association for Computing Machinery, 2009, pp. 55–68. [Online]. Available: https://doi.org/10.1145/1555816.1555823

[10] V. Issarny, V. Mallet, K. Nguyen, P.-G. Raverdy, F. Rebhi, and R. Ventura, "Dos and Don'ts in Mobile Phone Sensing Middleware: Learning from a Large-Scale Experiment," in *ACM/IFIP/USENIX Middleware 2016*, ser. Proceedings of the 2016 International Middleware Conference, Trento, Italy, Dec. 2016. [Online]. Available: https://hal.inria.fr/hal-01366610

[11] K. Lee, A. Agrawal, and A. Choudhary, "Mining social media streams to improve public health allergy surveillance," in *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015*, 2015, pp. 815–822.

[12] P. Tzamalis, P. Vikatos, and S. Nikoletseas, "A hybridization of mobile crowdsensing, twitter analytics, and sensor data for the holistic approach of pollen onsets detection," in *2019 15th International Conference on Distributed Computing in Sensor Systems (DCOSS)*. IEEE, 2019, pp. 188–191, poster paper.

[13] V. "Vlachos, Y. C. Stamatiou, P. Tzamalis, S. Nikoletseas, and K. Chantzi, "A social network analysis tool for uncovering cybersecurity threats," in *6th International Symposium for ICS & SCADA Cyber Security Research 2019 (ICS-CSR)*, Sep 2019.

[14] L. Klimek, K.-C. Bergmann, T. Biedermann, J. Bousquet, P. Hellings, K. Jung, H. Merk, H. Olze, W. Schlenter, P. Stock *et al.*, "Visual analogue scales (vas): Measuring instruments for the documentation of symptoms and therapy monitoring in cases of allergic rhinitis in everyday health care," *Allergo journal international*, vol. 26, no. 1, pp. 16–24, 2017.

[15] P. E. Taylor, R. C. Flagan, A. G. Miguel, R. Valenta, and M. M. Glovsky, "Birch pollen rupture and the release of aerosols of respirable allergens," *Clinical Experimental Allergy*, vol. 34, no. 10, p. 1591–1596, 2004.

[16] J. Ramos *et al.*, "Using tf-idf to determine word relevance in document queries," in *Proceedings of the first instructional conference on machine learning*, vol. 242, no. 1. Citeseer, 2003, pp. 29–48.

Appendix A
Security

Regarding the collection and storage of users' personal data to our infrastructure, a secure database with the appropriate authentication steps is used. In addition, platform identification credentials are uniquely encrypted using the Bcrypt library[22] provided by Flask micro web framework. Lastly, the SSL encryption protocol is used to securely exchange requests and responses between server and client, as well as a CSRF token to protect against relevant malicious attacks.

Appendix B
Additional Tables & Images

TABLE II
TERMS AND HASHTAGS USED IN QUERYING THE TWITTER STREAMING API.

Type	Terms
Keywords	pollen allergy grass, pollen health allergy, allergy pollen allergic rhinitis symptoms, allergic rhinitis symptom, suffer pollen allergy, pollen allergies, pollen symptoms, allergic pollen, trees allergy, flowers allergy, weeds allergy, molds allergy, pine allergy, allergic rhinitis disease
Hashtags	#allergic #rhinitis #disease, #allergic #rhinitis #symptom, #allergic #rhinitis #symptoms, #pollen #allergen, #pollen #allergens, #allergic #symptoms, #allergyseason #rhinitis, #pollen #rhinitis, #allergy #rhinitis, #pollen #spring #rhinitis, #pollen, #allergic #rhinitis #allergy

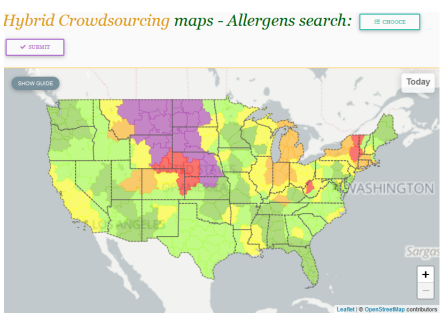

Fig. 9. Hybrid crowd-intelligence mining as derived from the Twitter and MCS components.

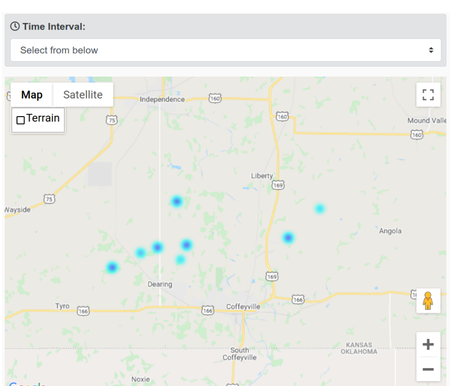

Fig. 13. myHealth system: The myAllergyMap feature.

[22]https://flask-bcrypt.readthedocs.io/en/latest/

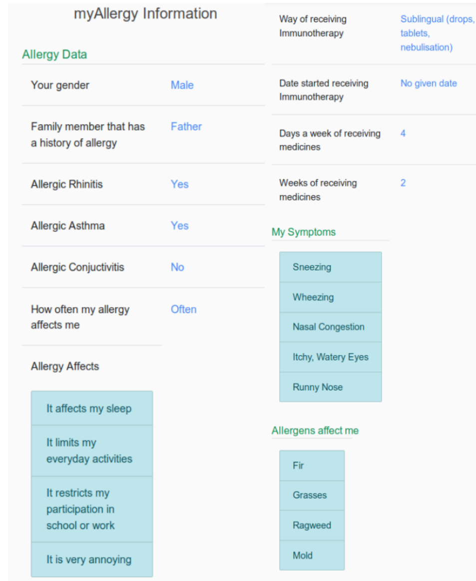

Fig. 10. User's personalized allergic rhinitis profile.

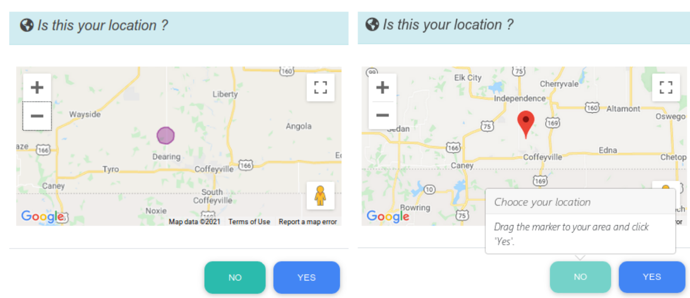

Fig. 11. The user interactive localization map, in case of both automatic spatial detection and non-available automated geolocation service where the geolocation recording is inserted manually.

Fig. 12. Time-series analysis of allergic-rhinitis-related tweets.