Sensor Data Harmonisation and Longitudinal Health Record Fusion: AI-Driven Approaches to Wearable-Clinical Data Integration for Personalised Healthcare

Dr. Siarhei Katsevich, Associate Professor of Computer Science, Belarusian State University of Informatics and Radioelectronics (BSUIR)

1. Introduction to Wearable Health Data Integration

The role of technology in influencing patient care is rapidly evolving. New and innovative methods for capturing health information in real time are appearing. One approach that shows considerable promise is the use of wearable health sensors to collect important health data. Typically focused on fitness and exercise applications, present sensor systems give consumers insights into a range of lifestyle and health factors. The potential of these devices and platforms is enhanced by their ability to collect real-time data that could be relevant to the management of chronic disease.

The seamless integration of wearable data with healthcare records, however, presents a number of challenges. First, the privacy and security of the data must be assured when interacting with an experimental system. Second, wearable data is collected according to the proprietary whims of the technology company producing the device: while there may be some effort toward tobacco control, there is no standardization in the components recorded, the format of that data, the protocols for data collection, or even the output that analyses provide. Consequently, to best use these technologies within the healthcare ecosystem, their data first must be deconstructed and then reconstructed in a format that can be integrated directly into a healthcare record system. Finally, many healthcare information systems do not support so-called 'unstructured' data such as text, audio, and video, which may be present in mHealth information. Artificial intelligence and machine learning hold significant promise to address these challenges related to data privacy, standardization, and turning data into information. They can accomplish this by generating a deeper understanding of health using more information and extracting actionable insights from that understanding.

1.1. Significance and Challenges

The idea of augmenting clinical and research investigations with data from people's wearable devices and smartphones is attracting increasing interest. Integration of wearable health data with clinical records promises more personalized, predictive, and preventive strategies for individual care or the management of broader populations. Timely point-of-care interventions are possible in response to trends in an individual's health status or lifestyle behaviors, leading to a paradigm shift in healthcare delivery. For research, continually acquired data can be used to measure participants in their real-world settings, leading to more objective quantification of health and disease and the potential to revolutionize clinical trials. The inclusion of wearable health data can also independently indicate health and disease risk or otherwise serve to sub-stratify individuals. However, there is still a wide gap between the potential of such technologies and their clinical impact, with many challenges impeding progress. There are technical and practical challenges in integrating wearable health data with health records, and the field is hampered by a number of practical, ethical, policy, and regulatory issues.

In terms of the data collected from wearables and smartphones, much of this is not collected to clinical grade. Inaccuracy and lack of calibration of the sensors also emphasize the need to support or combine this data with other emergent digital diagnostics and data sources such as electronic health records. Feasibility and interoperability need to be addressed across largely consumer-driven devices and healthcare systems, and there is a need for devices to include validation from regulatory bodies. Real-world use also emphasizes the issues of privacy and data security. It will be important that any systems put in place are sensitive to the potential for exploitation of individuals from misuse of their confidential and personal data. As with any database, there are numerous technical aspects to be considered, including the development of common fundamental data models facilitating data sharing across records and the comparability of the data generated. Exacerbating this, an additional difficulty that needs to be factored in is the widespread issue of data generated by consumer wearables or smartphone apps that are rarely integrated with any current electronic health records. Patent-related data in health records can also limit access. Finally, all of these issues raise an additional barrier: the legal and ethical issues also need to be addressed, as raising patient consent to enable their data to be shared and used for specific purposes could form the basis of a number of potential protections for innovators. It is a difficult area and a pressing issue. Most patients are becoming increasingly aware of the possibility of exchanging data for services, but the industry is currently not well-developed in this area, and thus regulation is lacking. Entering the space will require a great deal of care if any such governing data exist. Regulatory requirements that directly govern access to data are yet to be defined.

2. Machine Learning Techniques in Health Data Analysis

Machine learning approaches are powerful tools for extracting meaningful patterns and signals from highly complex, multi-faceted health datasets. They offer promise for bringing together wearable health data and analysis with electronic health records and other clinical descriptors and images. Fundamentally, machine learning approaches emphasize the ability to move beyond surface-level statistics to build tools that predict health outcomes and offer the potential for personalized care. For example, a decision model created by machine learning might take a set of physiological data and information and predict whether that patient is likely to develop a health condition in the future, and if so, to what degree.

The primary strength of machine learning approaches for processing health data arises from their ability to work with high-dimensional data where human vision might not be able to uncover patterns simply by examining images or raw datasets. The increasing interest and development of big data analytics and artificial intelligence with healthcare have bolstered machine learning's relevance and have shaped a dedicated field, healthcare analytics. The interest not only stems from the aspect of being able to process vast amounts of data but also in benefiting from the various algorithms that can assist in clinical decision support, diagnosis, risk stratification, patient-based outcome prediction, and the evolution of disease states throughout the clinical continuum. These algorithms can also perform administrative or financial analysis and automation associated with the evaluation of this data.

2.1. Supervised Learning Algorithms

Supervised learning is a subfield of machine learning that widely contributes to analyzing health data. It takes a dataset with input and output features and builds a model to make predictions. Features are variables from each data instance, and they help to determine the output. This model needs to be able to predict output variables from new input variables. However, model-driven selection is suitable when the target function can be expressed as a sought linear or non-linear function, which is not likely in medical cases. Various algorithms and models can fulfill tasks for classification, regression, and survival analysis. Especially in healthcare, supervised learning has been widely used to predict disease occurrence, diagnosis, prognosis, and patient treatment and response.

Many supervised learning approaches have been fruitfully applied to multimodal data in healthcare. Most commonly used algorithms include linear regression, decision trees, random forests, gradient boosting machines, support vector machines, naive Bayes, k-nearest neighbors, artificial neural networks, logistic regression, and ridge regression. Moreover, many deep learning models such as convolutional neural networks, recurrent neural networks, and long short-term memory may be utilized in patient data to predict mortality, diagnose diseases, and their progression paths. Supervised learning algorithms yield greater accuracy rates than unsupervised learning mainly because they use labeled data. Furthermore, supervised learning models can be more interpretable compared with unsupervised learning models. These advantages are balanced by disadvantages such as model overfitting and the need for abundant training data. Model performance can be enhanced by feature selection and quality data. The ability to make reliable predictions is closely aligned with these two points.

2.2. Unsupervised Learning Algorithms

While the vast majority of the alterations that have appeared in the literature are based on supervised learning algorithms, unsupervised learning is rapidly gaining interest, especially for the analysis of health data, as patients can be compared to each other based on large amounts of data without the need for any outcome markers, which can be time-consuming to collect. Unsupervised learning is known to be particularly rich in the health informatics literature, as it enables the hidden patterns to be illuminated that may otherwise remain difficult to see. While unsupervised learning is applied less in the context of clinical data, the same general principles can be applied in order to disaggregate large clinical data sources and build new understandings of the populations that those sources encapsulate. Unsupervised approaches such as clustering can reveal previously unrecognized associations between different diseases based on patterns of molecular biology and provide insights into how diseases can be categorized based on their pathological underpinnings. In this way, the information from unsupervised learning augments supervised learning, in which a model is built first and judgments are then made based on the supervisory marker.

Clustering methodologies, such as K-means, mixture models, and others, have been deployed extensively in order to segment patient populations. In addition to clustering techniques, unsupervised learning methods incorporating dimensionality reduction can be used to parse the complex relationships between items. Efforts have used these techniques to aggregate large multisource data in order to reduce the dimensionality and build low-dimensional representations of the overall puzzles and to incorporate that information into unsupervised cluster analyses. This use of structural prior information can be seen as a key approach that enables unsupervised learning to be used in the health context, where ever-evolving data characteristics are difficult to harness in some strong a priori manner. The development of formal ontology storage systems and standard terminologies is a related development that enables the facets of health to be described in explicitly linked terms. The background of unsupervised techniques is theoretical and varied. Some methods, such as K-means clustering, are simple procedures with clear drawbacks; others, such as mixture models, are more elegant methods, but they become more computationally expensive as their distributions become more complex. The primary advantage of unsupervised learning approaches is their flexibility to accommodate new data as they are produced. The practical disadvantage is that their findings can be difficult to interpret, as the result may consist of disease clusters that are difficult to parse without further biological content. Failure to approach these methods correctly can result in the discovery of spurious effects or relationships that are not grounded in any true biological or clinical meaning and can confound the discovery of meaningful relationships.

3. Integration of Wearable Health Data with Electronic Health Records (EHRs)

Wearable health sensors and devices have become popular for delivering health information about, for example, physical activity, heart rate, or glucose levels. Incorporating this information systematically with electronic health records for use in patient care has the potential to enhance patient accessibility, inform diagnosis and treatment, and evidence progress and outcomes. Challenges include the substantial heterogeneity in wearable devices, a lack of standardization, the need for validated and evidence-based data interpretation, appropriate alerting and automation strategies, and ensuring the security and privacy of data. A number of technological approaches can assist in data exchange. Many devices come with application programming interfaces that allow the sharing of device data with app or software developers. Some devices are compatible with well-established health information technology standards, and specialized data interoperability frameworks have been developed.

Advances in the digitalization of health enable the collection of large amounts of data about individuals' day-to-day function in a range of settings. However, health visibility is incomplete if data from wearable devices and other remote monitoring technologies are not systematically integrated with structured data held in electronic health records. Electronic health records are electronic, longitudinal records of patient health information generated by one or more encounters in any care delivery setting and include a range of data for patient care directly related to the current and previous clinical encounters.

3.1. Data Preprocessing and Cleaning Techniques

Before any analysis of data, there is a need for data preprocessing techniques, as the inclusion of faulty, incomplete, or redundant data may lead to misinterpretation and flawed decision-making. Research has shown that up to 80% of the effort in data analytics is likely to be on preprocessing. The typical data preprocessing techniques involve normalization, transformation, missing value treatment, outlier detection, and noise reduction—among many others. For processing healthcare data, data cleaning is the most significant stage as the correctness and reliability of any analysis are directly dependent on it. There are many tools and techniques for preprocessing when dealing with low-level health data. High-level health data, such as the residual of these carefully preprocessed raw sensor data, demand usual time series techniques for preprocessing, where equivalent components, low variability elements, noise, and periodic complex variations can be considered. In addition, generalities such as alert systems that notify nurses about reduced levels of patients' thirst for a consistent period may be derived. Systematically handling these systems to fit with high variability and large numbers of such time series streams is also a possible goal as part of preprocessing. There are other challenges in preprocessing healthcare-related data. Health data have value when analyzed in real-time, especially in chronic disease management while capturing the clinical fluctuations. Health-related data with intelligent real-time inference processing is a novel area. This means a need for streaming data processing, which is not a direct time series but at message levels.

4. Applications in Comprehensive Health Monitoring

Wearable health data, conveniently collected from various body locations, continuously and in real time, hold promise for comprehensive health monitoring. The data can deliver actionable insights into the health status of the individual or groups of individuals and can be used to trigger timely alerts and targeted interventions. Continuous connectivity and simultaneous data acquisition from multiple sensors in wearables allow monitoring of certain essential physiological parameters such as heart rate, activity, posture, sleep, and respiratory traces. Changes in these essential vital signs can alert about the changing health status of the patient. Wearable devices can simultaneously monitor multiple health parameters, and by applying AI-driven analytics, they can also provide comprehensive health assessments. In addition to key health indices, wearables are also capable of providing various other parameters, which can assist in the management of a range of chronic diseases.

For example, diabetes can be managed on the basis of real-time glucose data trends in the patient using a device. Real-time monitoring and real-time response strategies have continuously rolled out in the atmosphere. A closed loop of continuous glucose monitoring with embedded algorithmic recommendations and insulin delivery has already been approved for pediatric patients. AI and machine learning techniques are increasingly being applied to electronic health records and administrative health care data to identify high-risk individuals, offer patient-specific care and treatment, predict health outcomes, modify surgical interventions, design cost-effective service models, and improve the overall quality of patient health services, patient safety, and patient satisfaction. The AI-driven analysis of real-time continuous wearable data and wearable/phone audio and speech data will also integrate with the ongoing trends. In this role, wearables also promote the patient to be an informed citizen capable of having access to and learning about a proactive and connected health management monitoring system, consulting, making comparisons, and co-creating customized and personalized care. The consumers and the elite global population's active indulgence in preventive care signals a gradual convergence of 'casting and caring.' Patients and their 'willing' participation, when combined with passive and lifestyle data, and wearables and genomics, will be reshaped in the next decade as a major strength of health systems.

4.1. Real-Time Health Status Monitoring

There is a strong push in the healthcare research community towards real-time health status monitoring using wearable health data. Real-time descriptions allow for an immediate response, conducted as quickly as possible after the appearance of a certain sign, signal, symptom, or condition. One of the primary purposes of real-time monitoring for healthcare is to provide insight into the current status and conditions of a patient's health. In modern clinical settings, patients are often quickly monitored through vital signs and integrated into the overall clinical record in addition to focused tests. Regularly, patients are given check-ups to evaluate whether a particular treatment or procedure is effective.

The most apparent advantage of collecting data continuously is the capacity to gather additional data for a longer and defined period of time. Using this concentration of data collection, subtle changes can be detected in comparison to intermittent measurements for the same type of data. These subtle changes could be indicative of a warning of future problems. Wearables are capable of collecting real-time vitals, supporting real-time monitoring. The rapid changes in a patient's real-time health readings can often initiate a care pathway change. This can be immediately incorporated into the monitoring platform system and alert the relevant healthcare provider that the readings have changed. Several technological components are essential for real-time vital monitoring. These can include devices such as wearables responsible for data collection, integrated mobile applications supporting data display or input, and cloud services with temporary data storage or processing capabilities. Future internet-of-things approaches can use a direct interface from the wearable to a hospital EMR or any software supporting an API. The real potential level of data is dependent on the capabilities of the wearable device to transfer data in addition to the software employed. A challenge in real-time health monitoring includes not being inundated with an overabundance of information. The challenge can also be in discerning and interpreting the most appropriate information that is essential to an intuitive decision. The interpretation must abstract trends as the data is gathered, alert medical providers often or sometimes just in case of a crisis, and support future planning and prevention for patient care.

AI can be used to elevate the usefulness of real-time monitoring. AI can learn various patient data readings over time in addition to trends that potentially indicate a patient's health decline. For example, in a percentage of patients exhibiting certain data readings, a significant decline is indicated within a specified time frame. A clear warning could flash in the monitoring platform system as well as to a directly connected healthcare provider in a scenario when a patient is indicated with concerning readings. Such integrative approaches and new strategies can emphasize patient-centered care. A new paradigm in patient health and how we treat patients can be realized. Identifying when a patient is well or declining even before they start to feel ill will be indicated through subtle data trends instead of currently presenting vaccines or treatment once infection symptoms are noted. This is how we advance patient-centric care. This aspect integrates psychological and emotional cues such as mood in addition to other physiological or psychological patient indicators, allowing for overall patient wellness interventions to be applied.

5. Personalized Care and Treatment Plans

One of the primary goals of healthcare today is to provide care as well as form treatment plans for individual patients. Traditionally, care plans were based on the average patient size, the average disease manifestation, and the average health-related behavior. Over the last twenty-five years, we have, however, come to the realization that applying one-size-fits-all solutions to healthcare gives mediocre results. In order to significantly improve patient health, experts have thus started designing healthcare to provide not just diagnostic and treatment strategies, but also feedback to patients based on an accurate evaluation of their behavior and disease status in real time. This work is based in lifestyle medicine, which is an approach that uses interventions such as eating and exercise to affect patient health outcomes.

In one study, wearable data were used to perform a rigorous analysis of drug adherence in a patient population. Based on the findings, three possible stock keeping units of medication were suggested to the patient for further investigation, which is in line with medication being tailored to a patient's needs. It is now possible, because of developments in data analytics, to merge the data from a wearable with patient preference, disease state, socio-economic status, genomic integrity, and psychological state; in this way, treatment becomes the outcome of a personalized equation that we in healthcare try to optimize according to the patient's needs. The ultimate optimization in a chronic or lifestyle-related disease would be an ecological momentary intervention and momentary assessment. Simply put, the entire behavior matrix that a human has can change at any time point; hence, the care plan has to be devised at regular intervals. Lastly, if we want to change behavior through this concept of ecological momentary assessments and interventions, we need the compliance of the targeted patient group. We are no longer able to change people or study them once they are somewhat different from bio-specimens. This knowledge is a fundamental aspect of designing care plans for our patients.

5.1. Risk Stratification and Predictive Analytics

Often, the first vital step in making sense out of vast data pools is to group people with similar characteristics through risk stratification. Doing so allows interventive efforts to be focused more efficiently toward patient subgroups with higher risks. Building predictive models that integrate real-time data streams with the historical electronic health record in order to anticipate potential health events is one example of the use of machine learning in healthcare. These predictive models are used to suggest communication and treatment strategies. Machine learning techniques such as elastic net, random forest, generalized additive modeling, least absolute shrinkage and selection operator, gradient boosting machines, and support vector machines are used for predictive modeling. By involving historical data and real-time personalized monitoring data, these models assist timely treatment plans for patients and data-driven clinical decision-support tools. Furthermore, it can lead to long-term advantages such as better resource distribution and a reduction in unnecessary emergency department visits. The real potential of integrating patient wearable health data along with electronic health records is for managing resource misuse more effectively and driving preventive care.

Unfortunately, the ethical use of predictive analytics in healthcare is not straightforward, as it necessitates assessment of its broader implications such as potential bias, higher false alert rates, added workload, and its clinical utility. Machine learning models are known to be interpretable, biased, and unfair, and require large datasets for effective use. We can contemplate how continuous wearable patient data could be used for population health risk assessment. No doubt, retrofitting wearable health data into electronic health record analyses and patient management strategies will transform and augment risk assessment and preventive strategies. Machine learning models, predictive scoring systems, or risk stratification algorithms, when incorporating prevalent wearable data, can potentially diagnose and prevent accelerated health events among our population more effectively. As part of care coordination efforts, risk stratification can target patient groups for better population health management. By doing so, healthcare systems become more cost-effective and are prepared to manage emergent threats to populations.

6. Future Direction

The advances in AI and machine learning have revolutionized technology, and their application in integrating wearable technologies into clinical records is highly promising. Wearable technology has also been constantly evolving, with the launch of newer, sleeker devices with longer battery lives, capable of capturing and transmitting an increasingly large volume of health data. With health systems in various countries making their data more interoperable, it will be easier to integrate the data from a person's own device with their clinical records. Furthermore, social care practices, of which telehealth and remote monitoring will start to form a part, are also advancing in many countries. The scope of tasks covered by these services and the data used are also expected to expand, making such technological developments all the more pertinent. Challenges to be overcome in this effort will include ensuring that personal health data is secure, and that the ownership of this data is determined alongside appropriate regulatory frameworks. Moving forward, this form of technology may encourage a move towards a more integrated approach to patient empowerment, utilizing data-driven insights about the health of individuals and entire communities to improve patient welfare and reduce the burden on healthcare providers. The nature of personal health data may need to be revisited and the use of such data regulated at a policy level more closely, with interactions between healthcare policy, data usage, and ethical considerations made in order to update existing frameworks. The use of extensive personal health data might lead to more appropriate medical treatment for the patient, while regulatory restrictions on the use of this information might also need to be updated in order to reflect current approaches to data storage and analysis while protecting the rights of the individual.

7. Conclusion

Integration between wearable health data and clinical records is seen as providing a substantial value proposition for all stakeholders, not least for patient care simply by virtue of the fact that anything that provides more information and insights into what is happening with a patient does genuinely work towards personalisation of care and improved outcomes. Indeed, from the very top perspective of "medicine", it is often patient groups and rare and complex diseases that are used as exemplars as, being often highly heterogeneous in individual effects for individual patients, they benefit most from "smarter" diagnostic and therapeutic approaches. Clearly, off-the-shelf consumer wearable devices integrated into clinical systems have proliferated en masse globally, where in the US alone at least 35 examples can be found. The evidence for benefit sits alongside the commercial enthusiasm to integrate wearable device data into its patient data repositories and this is fostering substantial investments by healthcare IT companies especially in the Americas, Europe but also in other territories, and also giving rise to several clinical partnerships. However, and perhaps unsurprisingly those efforts buckle under the immediate and longstanding frustrations detailed earlier.

There remains a significant dislocation between wearable technologies and the needs of the health system. Why? We would argue that the points made which dwell on complexities, inadequacies and fears in several areas build the architecture of roadblocks which are all stacked on these issues regarding the foundational building blocks of a data-led continuum of care. If those could be addressed, it is not a giant leap to posit how wearable technology could evolve. The previous section has reported the incredible advances happening in "lab" research and the range of biomarkers that can be detected by various forms of technology integrated into wearables. Indeed, in the last year, it was reported that wrist-worn devices could detect possible long-term clinical conditions years in advance. This is the future proposition the sector and technology visionaries see for the integration of wearable data with clinical records. As such, the involvement of the health sector in trying to evolve this space to limit its challenges and maximise direct positive benefit to them will inevitably be alluring. It also means smaller niche markets could have their lives transformed by this type of advanced care based on data. So, despite all these barriers therein lies the potential for transformative value that can be drawn from more expansive bed-of-nails and types of data harvested by advanced analytics and self-learning algorithms.