For the application of machine learning in urinary incontinence, a systematic review

Nada OTMANI¹, Noura QARMICHE¹,², Mohammed OMARI¹, Imad CHAKRI¹, Hind Bourkhime¹, Soumaya BEN MAAMAR³, Ibtissam EL HARCH³, Bineta JHO DIAGNE³, Moncef MAIOUAK³, Nassiba BAHRA³, Nabil TACHFOUTI³ and Samira EL FAKIR³

¹Medical informatics unit, Laboratory of Epidemiology, Clinical Research and Community Health, FMPDF, Sidi Mohamed Ben Abdellah University, Fez, Morocco
²Laboratory Artificial Intelligence, Data Science and Emerging Systems, ENSA, Sidi Mohamed Ben Abdellah University, Fez, Morocco
³Laboratory of Epidemiology, Clinical Research and Community Health, FMPDF, Sidi Mohamed Ben Abdellah University, Fez, Morocco

ABSTRACT

Urinary incontinence (UI) is a common condition that can severely impact the person’s lives. Machine learning (ML) techniques are recognized as a promising method for help in decision making. This review aimed to summarize existing ML models applied to any aspect of this affection.

A systematic search was performed in PubMed, Scopus and Web of science. Original studies targeting urinary incontinence were included. The literature search identified 108 studies. 12 studies satisfied all the inclusion criteria and were retained. About 75% of them were for prediction. The most used machine learning’s algorithm to develop the UI’s models were SVM in three cases. The area under the curve (AUC), was reported for six of the models, and their values were between 0.64 and 0.87. The accuracy was reported by five models and varied between 0.775 and 1. The sensitivity and specificity are reported for only three models. Their values were respectively (0.72; 0.675); (0.804; 0.774); and (0.996; 1).

In this systematic review, many models showed improved discrimination and accuracy. More research is needed to continue promoting this field of research in ML in order to obtain solutions with multicenter clinical applicability before these models can be incorporated into routine clinical practice.

INTRODUCTION

Urinary incontinence (UI) is a common condition that can severely impact the lives of those who have it (Dunne. 2018) (Otmani. 2020). It was significantly associated with poor quality of life (QoL) (Pizzol. 2021) (Schurch. 2007) (Otmani. 2020). It has also a major impact on physical and psychological well-being of patients, and interpersonal relationships, especially for people with moderate to severe disabilities (Schurch. 2007) (Patrick et al. 2013). UI also represents a significant burden in terms of care (Ballanger 2005).

Artificial intelligence (AI) technologies are widely used in medical field. They have been developed to handle wide range of health data. Machine learning (ML) is one of the most common type of AI. Due to the large quantity and complex nature of medical information, their techniques are recognized as a promising method for supporting diagnosis or predicting clinical outcomes (Darcy, 2016) (Frizzell. 2017). It can help professionals in making decisions, reducing medical errors, improving accuracy in the interpretation of various diagnoses, and thereby reducing their workload (Makary, 2016).

In urinary incontinence, several ML models have been developed using different types of data, including electronic medical records, medical images, biochemical parameters, and biological markers. The type of data used depends on the MLojective (Werneburg. 2022) (Ahmed, 2020).

Objective

This present review aims to summarize existing machine learning models applied to any aspect of the UI.

METHODS

A systematic review, adhering to the recommendations of the Preferred Reporting Items for Systematic Reviews and Meta-
The Application of Machine Learning In Urinary Incontinence, A Systematic Review

Analyses (PRISMA) guidelines statement 2020 (« La déclaration PRISMA 2020: une ligne directrice mise à jour pour la notification des revues systématiques | Le réseau EQUATOR » s. d.), was performed to identify studies reporting the development or validation of a machine learning model for a prediction or a classification of any aspect of the urinary incontinence.

Search strategy and Information sources

We performed an exhaustive electronic literature search for English literature studies from January 2016 to September 2021. Literature was sought in Scopus, web of science, and PubMed.

The reference lists of studies selected for review were searched manually to identify additional potentially relevant studies. The search strategy was the following described in table 1:

Table 1 Search expressions used in the systematic review

<table>
<thead>
<tr>
<th>Database</th>
<th>Search expression</th>
<th>Year of publication</th>
</tr>
</thead>
<tbody>
<tr>
<td>Web Of Science</td>
<td>.“urinary incontinence” AND “machine learning”</td>
<td></td>
</tr>
</tbody>
</table>

Inclusion and exclusion criteria

The included papers adhere to the following selection criteria:

- Published as a primary research paper in a peer-reviewed journal;
- Describing the development and validation of a machine learning model for urinary incontinence
- The full article can be obtained in English

We excluded the systematic reviews, the meta-analysis, and bibliographic reviews.

One reviewer (N.O.) performed the search and screened the titles and abstracts to exclude papers that were clearly not relevant.

Two reviewers (N.Q. and S.E.) independently assessed a random selection of 5% of the papers each. The full text was examined when a definite decision to reject could not be made based on title and abstract alone. We discussed papers for which it was unclear whether or not the inclusion criteria were satisfied at a consensus meeting.

Data extraction

For each selected article, model’s data were extracted and analyzed. These data were about:

- General study characteristics, country, number of participants, target population, study design, year of publication, applied algorithms and retained ones,
- Model parameters, model performance including the AUC, accuracy, sensitivity and specificity

Data Synthesis, Visualization and Analysis

We used a qualitative, narrative synthesis method. The articles retained were described by their discrimination metrics. The accuracy was defined by Simundic (Simundic 2009) as the ability of a test to discriminate between the target condition and health. This discriminative potential can be quantified by several performance tools, such as sensitivity and specificity, AUC, accuracy metric, and other measurements.

To optimize the visualization of the results obtained in the systematic review, several tables and figures were made, showing mains characteristics of selected articles.

RESULT

Search results

The literature search identified 108 studies. After exclusion of duplicate studies from Scopus, web of science, and PubMed searches, and after stepwise exclusion of research outside the scope of our review, only twelve studies met all the inclusion criteria and were retained for further analysis. The PRISMA diagram for the systematic review process is shown in Fig1.

Models Development and Validation

Nine of the twelve models retained were for prediction, of which Five models were for the risk assessment of the UI. The three remaining ones were dedicated to classify the UI based on the bladder volume monitoring.

The machine learning algorithms used to develop the UI’s models were SVM in three cases (Dunne. 2018b) (Dunne. 2018a) (Santorelli. 2018), linear regression in two (Yahya. 2016) (Dunne. 2018), logistic regression in one (Omae. 2021), decision tree in one (Hung. 2019), random forest (RF) in one (Sheyn. 2019), an ensemble learning in two cases: a gradient boosting decision tree in one (Pan. 2020), and RF+ AdaBoost in the second (Jelovsek. 2018), and deep learning in two cases: ANN model in one (Sumitomo. 2020), and the deep learning survival analysis in the second (Trinh. 2021).

The table 2 & the figure 2 show the details of the development of each model.

Study Populations

Among all studies included, six used cohort studies to train the models (Jelovsek. 2018) (Hung. 2019) (Sumitomo. 2020) (Pan. 2020) (Omae. 2021) (Trinh. 2021). Four of them were based on image-based simulation data (Dunne. 2018b) (Dunne. 2018) (Dunne. 2018a) (Santorelli. 2018). One used medical record data (Sheyn. 2019), and the last one were based on a randomized trial study (Yahya. 2016).
Four studies were performed in the USA (Sheyn. 2019; Hung. 2019) (Pan. 2020) (Trinh. 2021), three of them in Ireland (Dunne. 2018b) (Dunne. 2018a) (Dunne. 2018), two in Japan (Omae. 2021) (Sumitomo. 2020), one in Malaysia (Yahya. 2016), one in Croatia (Santorelli. 2018), and one of them was a multicenter study (United Kingdom; New Zealand; and Swedish) (Jelovsek. 2018) (Table 2 & Fig 2).

Table 3 performance metrics used by articles

<table>
<thead>
<tr>
<th>Metric</th>
<th>Number of articles using this metric</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>AUC</td>
<td>6</td>
<td>50.0</td>
</tr>
<tr>
<td>Accuracy</td>
<td>5</td>
<td>41.7</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>2</td>
<td>16.7</td>
</tr>
<tr>
<td>Specificity</td>
<td>2</td>
<td>16.7</td>
</tr>
<tr>
<td>C-Index</td>
<td>3</td>
<td>25.0</td>
</tr>
<tr>
<td>Mean Absolute Error</td>
<td>1</td>
<td>8.3</td>
</tr>
<tr>
<td>Root mean squared error</td>
<td>1</td>
<td>8.3</td>
</tr>
</tbody>
</table>

Table 2 Details of the development of each model

<table>
<thead>
<tr>
<th>First auteur; year</th>
<th>Country of research; Year</th>
<th>Study description</th>
<th>Type of ML</th>
<th>Objectif of ML</th>
<th>ML algorithms</th>
<th>Performance metrics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yahya; 2016(Yahya et al. 2016)</td>
<td>Malaysia; 2016</td>
<td>Randomised trial (n=754; EBRT of the prostate to either 66, 70, or 74 Gy)</td>
<td>Prediction</td>
<td>Risk of UI</td>
<td>linear regression</td>
<td>AUC = 0.649</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Classification</td>
<td>Bladder volume monitoring</td>
<td>SVM</td>
<td>Sensitivity = 99.6; Specificity = 100; Accuracy = 100</td>
</tr>
<tr>
<td>Dunne; 2018(Dunne, Santorelli, McGinley, Leader, et al. 2018b)</td>
<td>Ireland; 2018</td>
<td>Image-based simulation data</td>
<td>Classification</td>
<td>Bladder volume monitoring</td>
<td>SVM</td>
<td>Accuracy = 100</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dunne; 2018(Dunne, Santorelli, McGinley, Leader, et al. 2018a)</td>
<td>Ireland; 2019</td>
<td>Image-based simulation data</td>
<td>Classification</td>
<td>Bladder volume monitoring</td>
<td>SVM</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jelovsek; 2018(Jelovsek et al. 2018)</td>
<td>United Kingdom; New Zealand; 2018</td>
<td>Cohort (2 cohorts: (1) women who gave birth in the United Kingdom and New Zealand (n=3783) and (2) women from the Swedish Medical Birth Register (n=4991))</td>
<td>Prediction</td>
<td>Risk of UI</td>
<td>RF+ AdaBoost</td>
<td>C-Index= 0.653, 0.689</td>
</tr>
<tr>
<td>Santorelli; 2018(Santorelli et al. 2018)</td>
<td>Croatia; 2019</td>
<td>Image-based simulation data</td>
<td>Classification</td>
<td>Bladder volume monitoring</td>
<td>SVM</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sheyn; 2019(Sheyn et al. 2019)</td>
<td>USA; 2019</td>
<td>Medical record (n=559; females )</td>
<td>Prediction</td>
<td>Response to treatment in UI</td>
<td>RF</td>
<td></td>
</tr>
<tr>
<td>Hung AJ; 2019(Hung et al. 2019)</td>
<td>USA; 2019</td>
<td>Cohort (n=100; )</td>
<td>Prediction</td>
<td>Risk of UI</td>
<td>DL</td>
<td></td>
</tr>
<tr>
<td>Sumitomo; 2020(Sumitomo et al. 2020)</td>
<td>Japan; 2020</td>
<td>Cohort (n=400; prostate cancer)</td>
<td>Prediction</td>
<td>Risk of UI</td>
<td>ANN</td>
<td></td>
</tr>
<tr>
<td>Pan; 2020(Pan et al. 2020)</td>
<td>USA; 2020</td>
<td>Cohort (n=86; patients who underwent prostate SBRT (40 Gy in 5 fractions)</td>
<td>Prediction</td>
<td>HRQOL changes in urinary function</td>
<td>gradient boosting decision tree</td>
<td>AUC = 0.87</td>
</tr>
<tr>
<td>Omae; 2021(Omae et al. 2021)</td>
<td>Japan; 2019</td>
<td>Cohort (n=630; community dwelling, independent older adults 75 years old or older who attended a health checkup in 2017 with a 1-year follow-up)</td>
<td>Prediction</td>
<td>UI as a predictor of falls of the elderly</td>
<td>logistic regression</td>
<td>Accuracy = 83.6; AUC = 0.818</td>
</tr>
<tr>
<td>Trinh; 2021(Trinh et al. 2021)</td>
<td>USA; 2019</td>
<td>Cohort (n=115; robot-assisted radical prostatectomy performed from July 2016 to December 2017)</td>
<td>Prediction</td>
<td>risk of UI</td>
<td>Deep learning survival analysis</td>
<td>C-Index= 0.708</td>
</tr>
</tbody>
</table>

**ML Models and Performance Metrics**

Discrimination, as measured by the area under the curve (AUC), was reported for 6 of the models, and their values were between 0.64 (Yahya. 2016) and 0.87 (Pan. 2020).

The accuracy was reported by five models and varied between 0.775 (Sumitomo. 2020) and 1 (Dunne. 2018a) (Dunne. 2018b).

Sensitivity and specificity are reported for only three models. Their values were respectively (0.72; 0.675) (Sumitomo. 2020); (0.804; 0.774) (Sheyn. 2019); and (0.996; 1) (Dunne. 2018b). The details of the performance metrics are shown in Table 3.

**DISCUSSION**

To the best of our knowledge, this is the first systematic review of the use of machine learning applied to urinary
incontinence. Even if the UI is a common health problem, our study shows that only twelve models are published in the last five years to predict or classify any aspect of urinary incontinence.

Given the heterogeneity of the models, we opted for a qualitative analysis of the results.

In this systematic review, several ML methods were used. The best discrimination model was reported by Pan (Pan. 2020) using logistic regression. The best accuracy was 100% found by Dunne et al in two studies conducted in 2018 (Dunne. 2018a)(Dunne. 2018b) and using each of them the SVM technique. It is important here to mention that it is difficult to define a best method for predicting urinary incontinence, because they differ in terms of input variables, aspect of urinary incontinence studied and sample size.

Except one (Jelovsek. 2018) of the retained studies, all of them had a sample size less than 1000 patients. Then the sample may not be representative for a given geographic group, representing one of the limitations of ML in health (Vayena. 2018).

The AUC was the metric most used (50%) in studies to describe the model performance. the comparison between models was impossible since the performance metrics used are different from one study to another. Thus, it is strongly recommended to unify the performance metrics of the different machine learning models (Simundić 2009).

The machine learning was used more for the UI risk prediction, or for bladder volume monitoring. However, some aspects of UI were less explored like the use of the ML in health related quality of life which was the subject of only one study.

The application of machine learning models in daily medical practice remains very low despite the recommendations of the majority of studies. This can be explained in part by the difficulty of using the models by physicians. The development of an easy-to-use framework for health care personnel is strongly recommended (Bertini et al. 2022).

CONCLUSION

In this systematic review, many models have shown good discrimination and accuracy. ML Models allow the identification of the most critical variables for clinicians, based on objective real-world data. However, it is necessary to continue to promote this area of ML research in order to obtain results with multi-center clinical applicability before these models can be integrated into routine clinical practice.

References


Omae, Kenji, Noriaki Kurita, Taro Takeshima, Toru Naganuma, Sei Takahashi, Takashi Yoshioka, Tsuyoshi Ohnishi, Fumihito Ito, Sugihiro Hamaguchi, et Shunichi Fukuhara. 2021. « Significance of Overactive Bladder as a Predictor of Falls in Community Dwelling Older Adults: 1-Year Followup of the Sukagawa Study ». The


How to cite this article:

********