

Systematic Review of Supervised Machine Learning Models in Prediction of Medical Conditions

Short title: Supervised Machine Learning Models in Prediction of Medical Conditions

Branimir Ljubic^{1,2}, Martin Pavlovski¹, Avrum Gillespie³, Daniel Rubin³, Galen Collier², Zoran Obradovic¹

1. Temple University, Center for Data Analytics and Biomedical Informatics (DABI), Philadelphia, PA 19122, U.S.A.
2. Rutgers University, Office of Advanced Research Computing (OARC), Piscataway, NJ 08854, U.S.A.
3. Lewis Katz School of Medicine, Temple University, Philadelphia PA 19140, U.S.A.

Corresponding author:

Zoran Obradovic

Mailing address:

Temple University, SERC 035-02
Center for Data Analytics and Biomedical Informatics
1925 N 12th St, Philadelphia, PA 19122

Phone number:

+215-204-6265

Email:

zoran.obradovic@temple.edu

ABSTRACT

37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Machine learning (ML) models for analyzing medical data are critical for both accelerating development of novel diagnostic and treatment strategies and improving the accuracy of medical care delivery. Our objective was to comprehensively review supervised ML models for diagnosis or treatment prediction. Publications indexed in PubMed were reviewed to identify articles utilizing supervised predictive ML models in medicine. Articles published between 01/01/2020–01/01/2022 were included in this review. Initially, PubMed was searched using MeSH major terms, and if more extensive search results were needed, a broader search was applied (titles/abstracts). PubMed indexed 21,268 published articles (MeSH Major topic) describing ML methods implemented in medicine. Of those, 11,726 articles were published within the last 2 years. Most of the published ML models in medicine in the last two years were different types of deep learning models (about 75%). Fifty articles were included in this review. Almost all categories of disease were subjects of ML predictions. Positive and negative factors in each of the scenarios need to be evaluated before the most optimal ML model is selected. Domain knowledge and collaborations between physicians and ML experts can improve the selection and prediction performance of ML models in medicine and facilitate implementation in clinical practice. Predictive ML models could provide recommendations to recruit suitable patients for clinical trials. Prediction ML models may contribute to development of more effective diagnostic and therapeutic choices, founded on evidence-based medicine. A broad range of methodological approaches have been taken toward this goal, and those approaches are presented here with their various advantages and disadvantages.

61

AUTHOR SUMMARY

62 Over the last decade, there has been rapid development of Machine learning (ML) methods to analyze
63 Big Data in medicine. ML is aimed to make the computer learn from past experiences and make
64 predictions by recognizing patterns in medical data. We performed a comprehensive systematic literature
65 review of recent publications (last two years), indexed in PubMed/MEDLINE that have described either
66 traditional or deep supervised prediction ML models in medicine. We identified 21,268 articles
67 describing ML implementation in medicine. 11,726 articles were published in the last 2 years. We
68 presented the number of publications describing each of the most often ML methods to show current
69 trends in development of these models. Most of the recently published ML models in medicine were
70 deep learning models. We found that the understanding of disease is likely to lead to more accurate
71 prediction. An important dilemma is the selection of optimal ML models for a specific task, considering
72 amount and type of available data. Domain knowledge and collaborations between physicians and ML
73 experts can improve the prediction performance of ML models, which could help clinicians to select the
74 most effective diagnostic and therapeutic choices available and decrease medical errors.

75

76

77

78 INTRODUCTION

79 Over the last decade, there has been significant growth of the amount of medical data generated by the
80 adoption and integration of electronic health records (EHR).[1] This growth in EHR data coincided with
81 rapid development of ML techniques and computing power to analyze Big Data in medicine, which
82 could contribute to improved medical solutions and better, more efficient healthcare.[2-5] ML is a branch
83 of Artificial intelligence (AI), aimed to make the computer learn from past experiences and make
84 predictions by recognizing patterns in medical data.[5-7] ML can be classified into three categories:

85 unsupervised, supervised, and reinforcement learning (RL). This paper focuses on supervised ML
86 techniques, where a function that maps an input to an output is inferred from labeled training data.
87 The objective of the research was to perform a comprehensive systematic literature review of recent
88 publications that have used either traditional or deep supervised prediction ML models in medicine. We
89 examined whether the method is appropriate for the selected medical prediction task, whether the model
90 is generalizable, and whether it could be used by clinicians to improve the quality of medical care.
91 Supervised prediction ML models are utilized in traditional and deep learning approaches.[5-7] The most
92 frequently used traditional ML models in medicine are: decision trees (DT),[8] random forest (RF),[9]
93 and other ensemble methods,[10-14] single and multi-layer perceptron (MLP),[15.16] Bayesian learning
94 (BL),[17] support vector machines (SVM),[18] k-nearest neighbors (k-NN),[19] linear regression
95 (LR),[20] and logistic regression (LogR).[21] Deep learning models are inspired by biological neural
96 networks, where each layer of the network learns higher order features of the previous layer. Different
97 types of neural networks have been designed, such as: deep neural networks (DNN) including deep belief
98 networks,[22] convolutional neural networks (CNN),[22] recurrent neural networks (RNN - long short-
99 term memory (LSTM) and gated rectified unit (GRU)),[23] etc.
100 Many ML models have been developed for prediction of diagnosis, or recommendation of the most
101 optimal therapeutic approach. Since this is a rapidly evolving scientific field, we reviewed articles
102 published within the last two years.

103

104 **MATERIALS AND METHODS**

105 We conducted a systematic review of articles describing supervised prediction ML models (traditional
106 and deep learning) published within the last two years (01/01/2020 –01/01/2022). The following
107 traditional supervised ML models were included in the review: DT, RF and other ensemble methods, the

108 perceptron, BL models, SVM, k-NN, LR, and LogR. Deep learning models included in the review were:
109 DNN, CNN, and RNN.

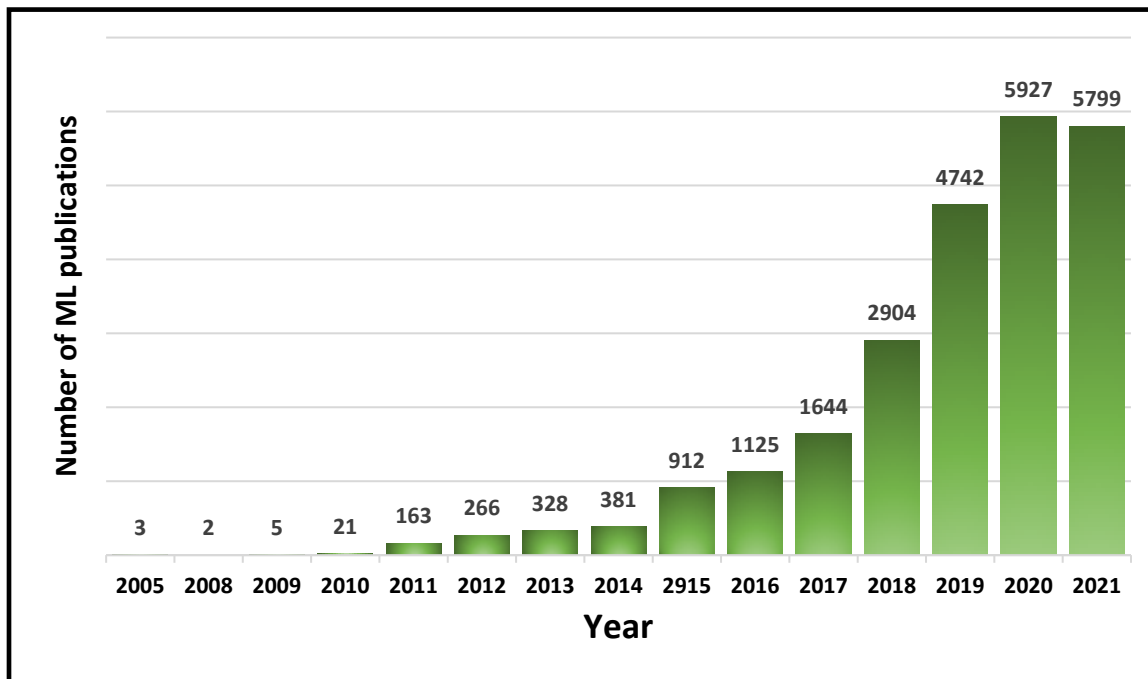
110 We searched PubMed/MEDLINE, as the most relevant source of medical topics, to identify articles
111 describing applications of supervised traditional and deep predictive ML models in medicine. Initially,
112 we searched PubMed using MeSH (Medical Subject Headings) major terms. If we did not find enough
113 literature for the specific ML model, we applied a broader search using terms that appear in titles of
114 publications, or in case of bagging and boosting ensemble models and RNN we used combinations of
115 MeSH and title/abstract searches. This searching approach extracted articles where the reviewed ML
116 model was a major part of the article. The search strategy used keywords indicating “ML model” AND
117 “prediction/detection/classification” AND “medical conditions (therapies, outcomes)”. In the example
118 of deep learning methods, we divided searches into 3 groups: classic DNN, CNN and RNN. To extract
119 general DNN models we used the following search query: ("deep learning"[Mesh] not "CNN" not
120 "RNN" not "LSTM" not "GRU"). For CNN models the query was: (CNN[Title]) OR (convolutional
121 neural networks[Title]). And for RNN model the query was: (RNN[Title/abstract] OR recurrent neural
122 networks[Title] OR LSTM[Title] OR GRU[Title] OR long short term memory[Title] OR gated rectified
123 unit[Title]).

124 We included articles that presented original research published in English. The search results were sorted
125 according to types of described ML models. For each of the reviewed ML models we selected a
126 representative sample. Priority was given to the newest published research in cases of multiple papers
127 describing similar predictive ML approaches. At least four authors agreed that the article was sufficiently
128 significant to be included in this review. We performed a systematic review of the literature to analyze
129 how useful and meaningful the described predictive ML models are in terms of realistic applicability in
130 medical practice.

131

132 **RESULTS**

133 We identified 21,268 articles (MeSH Major topic) describing ML implementation in medicine (Figure
 134 1). 11,726 articles were published in the last 2 years.



135

136

137 Figure 1. Number of publications describing ML applications in Medicine indexed in PubMed.

138

139

140

141 Fifty articles were included in the review. In Table 1, we present the method of search for each of the

142 ML models and the number of publications retrieved using that particular search method. Most of the

143 published ML models in medicine in the last two years were different types of deep learning models.

ML method	Type of ML	Type of PubMed Search	% of articles total	% in the last 2 years
Linear regression	Traditional	MeSH Major topic	13%	1%
SVM	Traditional	MeSH Major topic	13%	4%
DT	Traditional	MeSH Major topic	9%	1%
Logistic regression	Traditional	MeSH Major topic	9%	1%
Bayesian	Traditional	MeSH Major topic	5%	4%
RF	Traditional	MeSH Major topic	8%	8%
Ensemble - boosting	Traditional	Title search	1%	1%
k-NN	Traditional	Title search	3%	2%
Perceptron	Traditional	Title search	1%	1%
Ensemble - bagging	Traditional	Title search	1%	1%

DNN	Deep	MeSH Major topic	27%	35%
CNN	Deep	Title search	13%	22%
RNN	Deep	Title search	10%	14%

144

145 Table 1. Supervised ML methods applied in prediction of medical conditions. Percentages of articles of
146 the total number of publications by ML method are presented.

147

148

Traditional models

149

Decision trees are classification methods that adopt a top-down strategy where each node represents a
150 classification question and the branches partition the data into different classes.[8] DT models were
151 developed for prediction of diabetes mellitus type 2 (DM2), and essential hypertension (EH),[24] by
152 creating visually guided classification trees to facilitate the feature selection (four different datasets, sizes
153 547 – 12,447). The prediction accuracy of DM2 and EH in different scenarios varied between 58% and
154 87%. DTs predicted coronary artery disease with the accuracy about 91% on a dataset of 303
155 patients.[25] A DT algorithm was proposed to identify pre-treatment clinical predictors of survival in
156 rectal cancer (100 examples).[26] Predicted accuracy of survival rates were 71-76%. Presented DT
157 models have questionable generalization potential, since most of them were developed on small samples
158 frequently from one hospital.

159

Random Forest is a ML method that applies many decision trees to predict the outcome.[9] RF models
160 have been constructed to predict hepatotoxicity on a dataset of 346 samples, with the accuracy up to
161 71%.[27] This may provide a basis for improved safety evaluation in drug discovery and the risk
162 assessment of environmental pollutants. A retrospective review of 559 patients undergoing abdominal
163 hernia was the basis for RF modeling to predict surgical approach and determine the importance of
164 different socioeconomic variables in selecting the type of surgery (area under the receiver operating
165 characteristic (AUROC) ≈ 0.82).[28] Data were obtained from a single institution, which limits the
166 generalizability of findings. Psychotic and depressive symptom clusters in dementia were predicted
167 using RF on the EHR records of 4,003 patients with dementia (AUROC 0.80).[29] An RF model boosted

168 by the AdaBoost algorithm was utilized to predict the severity of COVID-19 cases and the possible
169 outcome, by using a patient's geographical, travel, health, and demographic data (3397 patients, accuracy
170 94%). The model revealed a positive correlation between patients' genders and deaths, finding that men
171 are more likely to die.[30] All presented RF models need further improvements including potential
172 combination with deep learning ML models to find a relevant application in practice.

173 **Ensemble Methods** (boosting, bagging) combine a few weak learning models and turn them into a great
174 learning algorithm.[10-14] The most famous boosting algorithm is AdaBoost. ML models were built to
175 predict the probability of different pairs of drugs and nanoparticles creating drug-decorated nanoparticle
176 (DDNPs) complexes with anti-glioblastoma activity. Forty-one features have been selected for 855,129
177 drug-nanoparticle complexes. The best model was obtained with the bagging ensemble classifier, based
178 on 20 decision trees with the AUROC of 0.96 and accuracy of 87%. This model could be applied for the
179 screening of nanoparticle-drug complexes in glioblastoma.[31] The bagging ensemble method improves
180 reproducibility of both cortical and subcortical functional parcellation of the human brain neuroimaging
181 (more than 300 samples).[32] AdaBoost was used to differentiate colorectal neoplasia from normal tissue
182 (AUROC up to 0.95 on 64 samples from 16 patients).[33] Both models use imaging data as inputs and
183 need additional research work, including comparison to deep learning models, before their potential
184 application in clinical practice.

185 **Perceptron and Multilayer Perceptron.** The perceptron is based on a threshold function that learns
186 weights for features and processes one example at a time. It could be used as a single-layer perceptron
187 or as a multilayer perceptron.[15,16] Modeling of the spread of the COVID-19 infection using a MLP
188 was designed on a dataset (20,706 examples) operated by the Johns Hopkins University Center for
189 Systems Science and Engineering (JHU CSSE).[34] This is one of many papers about COVID-19
190 intended to predict the spread of the infection, but the model lacks generalization ability. MLP was
191 applied in a diagnosis of breast cancer subtypes, using MRI images (704 images) with AUROC of

192 0.86.[35] The study used imaging data to distinguish between benign and malignant breast lesions. The
193 challenge remains how to effectively incorporate this model into everyday oncology.

194 **Bayesian ML** algorithms calculate probabilities for hypotheses. The class having maximum probability
195 is assigned as the most suitable class.[17] A study presented a method for classifying electrocardiogram
196 (ECG) data into four emotional states according to the stress levels using naive Bayes and SVM
197 algorithms with the average accuracy of the stress classification of 97.6%. Ability to quantify the stress
198 signals could facilitate a more effective management of mental state.[36] A Bayesian ML model was
199 designed to estimate the probability of an individual having an oral Human Papilloma Virus (HPV)
200 infection, given Oropharyngeal Squamous Cell Carcinomas (OPSCC) and other covariate
201 information.[37] The model is then inverted using Bayes' theorem to reverse the probability relationship.
202 The authors analyzed 8,106 OPSCC patients and achieved the AUROC of about 0.7. The Bayesian model
203 could be utilized to identify risk factors in estimating the probability of medical conditions.

204 **Support Vector Machine** models apply an optimization problem that attempts to find a separating
205 hyperplane with as large a margin as possible.[18] SVM was trained on 318 samples to distinguish
206 neurodegenerative movement disorders such as Parkinson's Disease (PD) from healthy subjects, and
207 from other movement disorders (precision \approx 81% and recall \approx 89%).[38] In this study, DNN and RF were
208 applied to the same task, with DNN achieving the best results. Models were trained on retrospective data
209 at a single site, with high data quality and none of the different classifiers outperformed the others, which
210 are some of limitations of these models. SVM models have also been used for diagnosis of early breast
211 cancer using PET images (116 samples, accuracy up to 85%, AUROC 0.89),[39] and detection of atrial
212 fibrillation (AF) using ECG data (79 AF and 336 non-AF cases, accuracy 97-100%).[40] Both models
213 use images as inputs. These models must be tested on bigger data and compared to deep learning since
214 recent literature show that deep learning models yield better prediction than SVM models on imaging
215 data.

216 **The K-Nearest Neighbors** algorithm takes into account k-neighboring points when classifying a data
217 point and assigns the class by finding the most prominent class among the k-nearest data points.[19]
218 Hepatocellular carcinoma (HCC) dataset from the UCI machine learning repository was used to test
219 different algorithms, including k-NN classifier, for feature selection and classification. The best results
220 achieved up to 84% accuracy.[41] The k-NN and SVM classifiers were utilized to determine whether
221 the patients have abnormal or normal respiration, or have bradypnea (slow breathing), or tachypnea (fast
222 breathing). The testing accuracies of the completely built SVM and k-NN classifiers were 96% and 99%,
223 respectively.[42] Constructed models must be tested on more datasets from other institutions to
224 determine reproducibility of the models. K-NN models have not been used very often for medical
225 predictions in the last few years.

226 **A Linear Regression** algorithm performs a regression task and predicts a specific value based on an
227 independent variable.[20] A linear model was created to predict the impact of the duration of exposure
228 (number of days) to COVID-19 on mortality rates (more than 270,000 patients).[43] Multiple regression
229 and LR analysis were successfully applied to predict the number of weekly deaths due to COVID-19 in
230 India (606 patients).[44] LR models are good options for time to event type of predictions and for
231 predictions of the exact numbers of patients, or the cost of care.

232 **Logistic Regression** uses the logistic function to binary classification and estimates the probability of
233 the event.[21] Multivariate LogR (as well as RF and XGBoost) models were applied for prognosis of
234 mortality risk in patients with COVID-19 (292 patients, AUROC \approx 0.95).[45] The model needs to be
235 tested on larger multi-center data. A study investigated the application of LogR and RNN LSTM models
236 in capturing clinical risk factors for outcome prediction of 575 patients with aneurysmal subarachnoid
237 hemorrhage (AUROC 0.89).[46] Since, the LSTM RNN model achieved higher accuracy, it is likely a
238 better choice in this type of study. A logistic regression-based ML prognostic algorithm is implemented
239 in real-time as a clinical decision support (CDS) system to facilitate decision making for patients with

240 suspected COVID-19 in the emergency department (ED). Training data included 1,469 adult patients
241 who tested positive for Severe Acute Respiratory Syndrome (SARS) within 14 days of acute care. The
242 algorithm performed well with an AUROC of 0.85.[47] A LogR based ML-enabled CDS can be
243 developed, validated, and implemented with high performance across multiple hospitals while being
244 equitable and maintaining performance in real-time validation.

245 **Deep learning**

246 **Deep Neural Networks.** DNN is a multilayer neural network with an input layer, hidden layers, and an
247 output layer. DNN learns weights so the output from the network correctly classifies the example. The
248 back propagation algorithm is a standard approach to train DNNs.[22] A compartmental model enhanced
249 with deep learning methodology predicted the dynamics of the COVID-19 epidemic in the U.S. using
250 the JHU CSSE data repository.[48] The model predicted the number of active cases between 3.2-3.3
251 million on August 16-18, 2020. The actual number of infected cases on August 16-18, 2020, was about
252 2.5 million (CDC data), so the model was not accurate. DNN methods were designed, to predict the
253 dynamics of the COVID-19 pandemic outbreak on JHU CSSE data and Korea Centers for Disease
254 Control and Prevention data.[49] Predictions of different aspects of COVID-19 epidemic are popular,
255 but considering the current state of the pandemic, it is difficult to confirm that these models work
256 accurately in reality. A DNN was designed for prediction of behavior of engineered RNA elements
257 capable of detecting small molecules, proteins, and nucleic acids.[50] This work shows that DNN
258 approaches could be used for predictions in RNA synthetic biology, but more data are needed for the
259 training of DNNs, as well as improvement of DNN architectures. A DNN performed an automatic
260 diagnosis of the 12-lead ECG recordings and outperformed cardiology residents in recognizing six types
261 of abnormalities, with F1 scores above 80% and specificity over 99%.[51] Additional studies could test
262 whether DNN effectively diagnose different ECG abnormalities, including myocardial infarction. DNN,
263 RF, and a simple statistical test were used to predict COVID-19 infections from full blood counts only

264 (598 samples), without knowing the history of the patients (accuracy up to 91%).[52] It appears that
265 these DNN models are more of a theoretical work at this stage of development, without evident clinical
266 implementation. A deep ML model was developed for automatic detection of brain metastases that uses
267 contrast-enhanced and non-enhanced CT images as inputs. The dataset contained CT scans of 116
268 patients with brain metastases (total of 659 metastases). Single-shot detector (SSD) ML models were
269 constructed with a feature fusion module. The sensitivity was 88.7% for the model that used both
270 contrast-enhanced and non-enhanced CT images (the CE + NECT model) and 87.6% for the model that
271 used only contrast-enhanced CT images (the CECT model).[53] The model is a contribution to imaging-
272 based diagnostics, and it needs further testing on larger datasets to improve generalization. Another study
273 proposed a novel deep learning architecture involving combinations of CNN layers and RNN layers that
274 can be used to perform segmentation and classification of five cardiac rhythms based on ECG recordings.
275 The algorithm is developed in a sequence to sequence setting where the input is a sequence of five second
276 ECG signal sliding windows and the output is a sequence of cardiac rhythm labels. Experimental result
277 shows this approach can achieve an average F1 scores of 0.89.[54]

278 **Convolutional Neural Networks** use a special kind of linear mathematical operation called convolution.
279 The hidden layers of a CNN typically consist of a series of convolutional layers.[22] CNNs showed great
280 potential for melanoma subtypes and localization diagnosis on dermoscopic image datasets (780 images)
281 and achieved AUROC ≈ 0.93 . [55] Improvements in the accuracy of this model could be achieved by
282 adding more training images of mucosal and subungual sites. Data-augmentation deep models
283 (DADLM) that enhance the learnability of CNNs and Convolutional LSTM (ConvLSTM) deep learning
284 models, improve the accuracy of COVID-19 detection.[56] The study used 50 images (X-ray and CT).
285 This model needs more reliable data to confirm its performance. The fast-track COVID-19 classification
286 network (FCONet) was developed to diagnose COVID-19 pneumonia in CT images (3,993) and
287 differentiate it from non-COVID-19 pneumonia and non-pneumonia diseases with $\approx 99\%$ accuracy.[57]

288 A CNN was also utilized to classify solid, lipid-poor, contrast enhancing renal masses using enhanced
289 CT images (143 patients) with the accuracy $\approx 99\%$, and AUROC ≈ 0.82 , [58] and for automated prediction
290 of breast cancer risk on 92 histopathological images where the F1 score was 0.73. [59] These CNN
291 models are examples of deep learning that rely on medical imaging. More research that uses EHR, or
292 medical text data in addition to imaging could contribute to faster and cheaper diagnostics. Transfer
293 learning is applied to train ResNet-50 and ResNet-101 deep learning models on augmented HAM10000
294 datasets, which contained about 42,000 dermoscopy skin cancer images. Achieved accuracy was better
295 when used augmented dataset compared to the original dataset and it was about 91.7%. [60] CNN model
296 has shown great success in decoding motor preparation of upper limbs from time-frequency maps of
297 EEG signals. [61] A deep learning architecture was applied to early diagnosis of glaucoma (301 images,
298 AUROC 0.92), [62] and for early diabetic retinopathy detection, [63] on retinal fundus images (40 images,
299 AUROC 0.94). Further studies with larger datasets, adding post-processing methods, and improved
300 optimized deep ML architectures could increase the accuracy of these models. An interpretable
301 classification approach of ultrasound images for the risk assessment and stratification of patients with
302 carotid atheromatous plaque was designed using CNNs and achieved AUROC of 0.73. [64] The
303 integration of interpretability methods with deep learning strategies can facilitate the identification of
304 ultrasound image biomarkers for the stratification of patients with carotid atheromatous plaque. A dataset
305 of 1,900 chest X-ray images has been used with the proposed CNN based model: “C19D-Net” to detect
306 COVID-19 with the accuracy of 96.24%. [65] The idea is to employ the constructed CNN ML model to
307 help radiologists improve their accuracy of detection of COVID-19 from X-ray images.

308 **Recurrent Neural Networks** are a type of neural networks that allow analysis of temporal heterogeneous
309 medical data. LSTM or GRU units effectively model the irregular visiting patterns in the long,
310 heterogeneous sequence of events in EHR. [23] RNNs and the magnetic induction system were integrated
311 to detect a wide range of human motions. [66] The benefit of LSTM RNN for sequence classification is

312 the ability to support multiple parallel temporal input data from different sensor modalities.[66] LSTM
313 and GRU RNN models were developed to predict complications of DM2 (two million patients with DM2
314 diagnosis), with the prediction accuracy up to 84%. They outperformed traditional ML models in
315 prediction accuracy of 10 selected complications of DM2.[67] An RNN approach was used for predicting
316 hemoglobin levels in patients with end-stage renal disease (7,739 patients) and produced mean absolute
317 error (MAE) of 0.54.[68] Further research is needed to incorporate the dialysis and laboratory
318 information. RNNs were designed for monitoring of depth of anesthesia based on features of EEG signals
319 (20 patients).[69] LSTM RNN models predicted AD from conditions, measurement, and drugs domain
320 on about 2,600 patients. A successful application of the drugs domain in prediction of AD was presented
321 (area under the precision recall curve (AUPRC) 0.99).[70] Additional research with the drugs domain is
322 required to develop comprehensive clinically applicable ML solutions. Another study utilized arterial
323 waveforms recorded on 18,813 patients during noncardiac surgery to predict short-term intraoperative
324 hypotension. A weighted average hybrid of deep learning CNN and RNN models performed the best
325 (AUPRC 0.716).[71] Intraoperative hypotension has an adverse impact on postoperative outcomes and
326 accurate prediction could improve survival. Acute kidney injury (AKI) is associated with poor patient
327 outcomes and increased health care costs. Two RNN algorithms were created using a dataset of more
328 than 72,000 patients.[72] Model 1 predicted the occurrence of AKI within 7 days with AUROC of 0.84
329 and model 2 predicted the future trajectory of creatinine values up to 72 hours with AUROC of 0.9.
330 Further development of the suggested approaches could incorporate the model into CDS systems for
331 prediction of in-hospital AKI.[72] Researchers leveraged a big dataset (48,151 patients) to build an RNN
332 to predict the risk of developing hepatocellular carcinoma (HCC). RNN models achieved AUROC of
333 0.759.[73] Deep learning RNN models outperformed traditional models, suggesting that RNN models
334 could be used to identify high risk of developing HCC. The performance of the presented RNN models

335 could be improved by training them on better quality hospital datasets and further optimization of deep
 336 learning models. We summarized the reviewed ML models in Table 2.

337
 338

ML Method	Author and year	Topic	
DT	Soguero-Ruiz C, et al 2020	Diabetes Mellitus, Hypertension	24
DT, SVM	Joloudari JH, et al 2020	Coronary artery disease	25
DT	De Felice F, et al 2020	Rectal cancer	26
RF	Chavan S, et al 2020	Liver toxicity	27
RF	Tracy BM, et al 2020	Hernia repair	28
RF	Mar J, et al 2020	Dementia, Neuropsychiatric symptoms	29
RF, AdaBoost	Iwendi C, et al 2020	COVID-19	30
Ensemble bagging	Munteanu CR, et al 2021	Glioblastoma, drugs	31
Ensemble, bagging	Nikolaidis A, et al 2020	Functional parcellation of the human brain	32
Ensemble, AdaBoost	Li S, et al 2020	Colorectal cancer	33
MLP	Car Z, et al 2020	COVID-19	34
MLP	Leithner D, et al 2020	Breast cancer	35
Bayesian, SVM	Kang M, et al 2021	Electrocardiogram signals, Mental stress	36
Bayesian	Tewari P, et al 2021	Oropharyngeal cancer	37
SVM, DNN	Varghese J, et al 2020	Movement Disorders	38
SVM	Satoh Y, et al 2020	Breast cancer	39
SVM	Lown M, et al 2020	Atrial fibrillation	40
K-NN, Boosting, SVM, DT	Angelis I, et al 2021	Hepatocellular cancer	41
K-NN, SVM	Jagadev P, et al 2020	Respiration rate (breathing)	42
LR	Verma V, et al 2020	COVID-19	43
LR, Multiple regression	Ghosal S, et al 2020	COVID-19	44
LogR	Ma X, et al 2020	COVID-19	45
LogR	Tabaie A., et al 2020	Aneurysmal Subarachnoid Hemorrhage	46
LogR	Lupei MI, et al 2022	COVID-19	47
DNN	Deng Q 2020	COVID-19	48
DNN	Jung SY. Et al 2020	COVID-19	49
DNN	Angenent-Mari NM, et al 2020	Genetics	50
DNN	Ribeiro AH, et al 2020	ECG diagnosis, Heart diseases	51
DNN, RF	Banerjee A, et al 2020	COVID-19	52
DNN	Takao H, et al 2021	Brain metastases	53
DNN, CNN, RNN	Pokaparakarn T, et al 2021	ECG Cardiac Rhythm	54
CNN	Winkler JK, et al 2020	Melanoma	55
CNN, DNN, LSTM	Sedik A, et al 2020	COVID-19	56
CNN, DNN	Ko H, et al 2020	Chest CT Image, Pneumonia	57
CNN, DNN	Oberai A, et al 2020	Renal tumor, CT scan	58
CNN, DNN	Wetstein SC, et al 2020	Breast tumors	59
CNN	Arshad M, et al 2021	Skin cancer	60
CNN	Mammone N, et al 2020	EEG signals, Motor upper limb	61
CNN	Muramatsu C. 2020	Glaucoma, Retinal fundus images	62
CNN, DNN	Hatanaka Y. 2020	Retinopathy	63
CNN	Ganitidis T, et al 2021	Carotid artery stenosis, Ultrasound	64
CNN	Kaur P, et al 2021	COVID-19	65
RNN	Golestani N, et al 2020	Human activity recognition	66
RNN, RF, MLP	Ljubic B, et al 2020	Diabetes mellitus, complications	67
RNN	Lobo B,	End-Stage Renal Disease, Hemoglobin	68
RNN	Li R, et al. 2020	Monitoring Depth of Anesthesia	69
RNN LSTM	Ljubic B, et al 2020	Alzheimer's disease	70

RNN	Choe S, et al 2021	Hypotension (intraoperative)	71
RNN	Kim K, et al 2021	Acute Kidney Injury	72
RNN	Ioannou GN, et al 2020	Hepatocellular cancer	73

339

340 Table 2. Reviewed articles, classified by ML model types. First authors, publication year, and medical
341 topics described in the publications.

342

343

344 The number of articles by types of medical conditions and topics is presented in Table 3.

345

Types of diseases by systems or therapy	Number of reviewed articles
Neoplasms (Oncology)	13
Nervous System and Sense Organs	12
COVID-19	11
Circulatory System	10
Endocrine, Nutritional and Metabolic	2
Respiratory System	2
Genitourinary System	2
Mental Disorders	2
Digestive System	1
Surgical procedures	1
Genomics, proteomics	1
New drugs, drug therapy	1

346

347 Table 3. The number of analyzed articles by types of medical conditions.

348

349 **DISCUSSION**

350 This review shows that most ML models in medicine represent great software solutions with high
351 prediction accuracy, but only handful of models could find an implementation in medical practice.

352 Neurological conditions are among the most common medical system subject to ML model
353 applications.[29,31,32,38,46,53,61-63,66,69,70] The most frequent type of data used in these

354 applications were imaging data. Images consist of spatially coherent pixels in a local region, meaning
355 that pixels close to each other share similar information. Deep learning architectures (especially CNN)

356 produce higher accuracy predictions from image inputs than from EHR type of datasets, which are often
357 heterogeneous. Another medical discipline extensively used in ML analysis is oncology. Accurate

358 predicting the development of cancers or complications of cancers could indicate earlier diagnosis and
359 therapeutic approaches that would improve outcomes. [26,31,33,35,37,39,41,53,55,58-60,73] The
360 majority of these ML applications use imaging data (most often histologic type) for classification of
361 malignant versus benign tumors. Cardiovascular conditions and DM are among the most common
362 medical conditions used in predictive analysis. [24,25,36,40,46,51,54,64,67,71] The challenges with
363 these types of predictions are often related to limitations of data availability. Insurance claims data was
364 frequently used but often lacks important clinical information such as laboratory results and
365 medications.[67,70]

366 Many traditional and deep ML models were utilized with the goal of helping to detect COVID-19
367 infections, complications, or outcomes as one of the most frequent research topic in the last two years.
368 [30,34,43-45,47-49,51,52,56,65]

369 The performance of predictive ML models in medicine depends on multiple factors. For challenging
370 prediction problems, the understanding of disease is likely to lead to more accurate prediction. Physicians
371 must be better motivated to use ML developments, which is not always easy to achieve since they
372 perceive this activity as something that decreases their time with patients.[1] Since many physicians use
373 computers daily, better presented benefits of ML prediction models could increase their adoption in
374 medicine. Evidence-based medicine requires statistical analysis of medical data and ML is a form of that
375 analysis. Some form of ML should become a part of statistics teaching in medical school to prepare
376 future physicians for meaningful adoption of medical ML models. To make ML more meaningful in
377 clinical practice, we should focus on tasks that physicians need help with and where results of ML could
378 help physicians to improve their decisions. The computers that physicians use for EHR could also be
379 used for ML models. Additionally, ML is relatively inexpensive compared to basic science and large-
380 scale clinical research.

381 Traditional ML methods do not always achieve accuracy that would convince medical doctors of the
382 benefits of the proposed predictive models.[24-47] A prediction accuracy of 70-90% is generally a good
383 result in terms of performances of ML models but may not be high enough to suggest clinically
384 meaningful improvements in practice. Traditional models have the advantage of simplicity and
385 interpretability but suffer from somewhat worse accuracy.[38,52,67]

386 The most successful and meaningful application of deep learning ML models was achieved in the
387 imaging field.[53,55-65] Analyses of CT scans, X-rays, Doppler ultrasound, histo-pathological images
388 obtained high accuracy results, which often outperform medical experts. RNN models capture the
389 temporal nature of EHR, imaging and other medical data to predict diseases, complications, and
390 outcomes.[66-73] Deep learning models produce higher accuracy but suffer from issues of
391 interpretability and instability.[15,75] Combinations of traditional and deep learning models could
392 address challenges of interpretability and accuracy.[38,52,67] Many datasets are small and do not have
393 enough samples for implementation of deep learning models. In those cases, traditional ML models are
394 the only option.

395 To build effective ML models, we must understand how to select relevant features to train ML models.
396 Computational methods that use optimization function to automatically select useful features have been
397 developed.[76,77] In addition to automatically selected features, we often use medical domain
398 knowledge to identify useful features that could help in improvement of predictions.[67,70] If analyses
399 point toward certain features as the most important for obtaining the model performance, the next
400 challenge is how to quantify the relevance of those features.

401 An important dilemma is the selection of the most optimal ML models for a specific task. We need to
402 consider which specific medical problem we want to solve. We must also determine how much and what
403 type of data are available, how much data are missing, and whether temporal information is included.
404 Implementation of ML in prediction of medical conditions using EHRs and other non-imaging data as

405 cheaper source of data could achieve meaningful results at lower cost. We would need to weigh positive
406 and negative factors in each of the options before we select the most optimal model for the given task.
407 Predictive ML models could potentially help to build CDS systems to make better medical decisions.
408 These models can provide recommendations to select suitable patients for clinical trials.[70] Domain
409 knowledge and collaborations between physicians and ML experts can improve the prediction
410 performance of ML models in medicine and facilitate implementation in clinical practice. Prediction ML
411 models could help clinicians to select the most effective diagnostic and therapeutic choices available.
412 Successful ML models can make medicine more efficient, improve outcomes, and decrease medical
413 errors. We predict that ML models will continue to develop, and they will be applied more broadly in
414 clinical practice.

415 416 **ACKNOWLEDGMENTS**

417 This research was supported by the National Institute of Diabetes and Digestive and Kidney Diseases of
418 the National Institutes of Health under Award Number R01DK122073. The content is solely the
419 responsibility of the authors and does not necessarily represent the official views of the National
420 Institutes of Health. The research was also supported by Clinical and Translational Science Award
421 (CTSA) Program grants under Award numbers: UL1TR003017, KL2TR003018 and TL1TR003019.

422 **AUTHORS CONTRIBUTORSHIP**

423 BL, MP, AG, DR, GC, and Z.O. performed the literature review and wrote the paper.

424 The authors have no competing interests to declare.

425

426 **REFERENCES**

427 1. Hoyt RE, Yoshihashi AK. Health Informatics: Practical Guide for Healthcare and Information
428 Technology Professionals, Sixth Edition, Morrisville, PA: Lulu Press 2014.

- 429 2. Gligorijevic D, Stojanovic J, Djuric N, et al. Large-scale discovery of disease-disease and disease-
430 gene associations. *Scientific reports* 2016;6(1):1-2.
- 431 3. Gligorijevic D, Stojanovic J, Satz W, et al. Deep attention model for triage of emergency department
432 patients. In *Proceedings of the 2018 SIAM International Conference on Data Mining*, pp 297-305, 2018.
- 433 4. Stojanovic J, Gligorijevic D, Radosavljevic V, et al. Modeling healthcare quality via compact
434 representations of electronic health records. *IEEE/ACM transactions on computational biology and*
435 *bioinformatics* 2016;14(3):545-54.
- 436 5. Waringa J, Lindvall C, Umeton R. Automated machine learning: Review of the state-of-the-art and
437 opportunities for healthcare. <https://doi.org/10.1016/j.artmed.2020.101822>.
- 438 6. Hamet P, Tremblay J. Artificial intelligence in medicine. *Metabolism* 2017;69:36-40.
439 <https://doi.org/10.1016/j.metabol.2017.01.011>.
- 440 7. Daume H. *A Course in Machine Learning*. Second edition. 2017. [http://ciml.info/dl/v0_99/ciml-](http://ciml.info/dl/v0_99/ciml-v0_99-all.pdf)
441 [v0_99-all.pdf](http://ciml.info/dl/v0_99/ciml-v0_99-all.pdf).
- 442 8. Breiman L, Friedman J, Stone CJ, et al. *Classification and regression trees*. Boca Raton, FL: CRC
443 press 1984.
- 444 9. Breiman L. Random forests. *Machine learning* 2001;45(1):5-32.
- 445 10. Breiman L. Bagging predictors. *Machine learning* 1996;24(2):123-40.
- 446 11. Schapire RE, Freund Y. *Boosting: Foundations and algorithms*. Cambridge, MA: MIT Press 2014.
- 447 12. Arsov N, Pavlovski M, Basnarkov L, et al. Generating highly accurate prediction hypotheses through
448 collaborative ensemble learning. *Scientific reports* 2017;7:44649.
- 449 13. Pavlovski M, Zhou F, Stojkovic I, et al. Adaptive skip-train structured regression for temporal
450 networks. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*
451 2017 Sep 18, pp. 305-21. Springer, Cham.

- 452 14. Pavlovski M, Zhou F, Arsov N, et al. Generalization-Aware Structured Regression towards
453 Balancing Bias and Variance. In IJCAI 2018 Jul 13, pp. 2616-22.
- 454 15. The perceptron. A Course in Machine Learning. Second edition. 2017.
455 http://ciml.info/dl/v0_99/ciml-v0_99-ch04.pdf
- 456 16. Rumelhart DE, Geoffrey EH, Williams RJ. "Learning Internal Representations by Error
457 Propagation". David E. Rumelhart, James L. McClelland, and the PDP research group. (editors), Parallel
458 distributed processing: Explorations in the microstructure of cognition, Volume 1: Foundation.
459 Cambridge, MA: MIT Press 1986.
- 460 17. Barber D. Bayesian reasoning and machine learning. Cambridge, MA: Cambridge University Press
461 2012.
- 462 18. Cortes C, Vapnik V. Support-vector networks. Machine learning 1995;20(3):273-97.
- 463 19. Cover T, Hart P. Nearest neighbor pattern classification[J]. Information Theory, IEEE Trans Inf
464 Theory 1967;13(1):21-7.
- 465 20. Weisberg S. Applied linear regression. Hoboken, NJ: John Wiley & Sons 2005
- 466 21. Murphy K. Logistic regression. Machine Learning: A Probabilistic Perspective, Chapter 8, pp. 245
467 -279. Cambridge, MA: MIT Press 2012.
- 468 22. Goodfellow I, Bengio Y, Courville A. Deep learning. Cambridge, MA: MIT press 2016.
- 469 23. Williams RJ, Zipser D. A learning algorithm for continually running fully recurrent neural networks.
470 Neural Comput 1989;1(2):270-80.
- 471 24. Soguero-Ruiz C, Mora-Jiménez I, Mohedano-Munoz MA, et al. Visually guided classification trees
472 for analyzing chronic patients. BMC Bioinformatics 2020;21(Suppl 2):92. doi:10.1186/s12859-020-
473 3359-3

- 474 25. Joloudari JH, Joloudari EH, Saadatfar H, et al. Coronary Artery Disease Diagnosis; Ranking the
475 Significant Features Using a Random Trees Model. *Int J Environ Res Public Health* 2020;17(3):731.
476 doi: 10.3390/ijerph17030731.
- 477 26. De Felice F, Crocetti D, Parisi M, et al. Decision tree algorithm in locally advanced rectal cancer: an
478 example of over-interpretation and misuse of a machine learning approach. *J Cancer Res Clin Oncol*
479 2020;146(3):761-5. doi:10.1007/s00432-019-03102-y
- 480 27. Chavan S, Scherbak N, Engwall M, et al. Predicting Chemical-Induced Liver Toxicity Using High-
481 Content Imaging Phenotypes and Chemical Descriptors: A Random Forest Approach. *Chem Res Toxicol*
482 2020;33(9):2261-75. doi:10.1021/acs.chemrestox.9b00459
- 483 28. Tracy BM, Finnegan TM, Smith RN, et al. Random forest modeling using socioeconomic distress
484 predicts hernia repair approach. *Surg Endosc* 2020. doi:10.1007/s00464-020-07860-6
- 485 29. Mar J, Gorostiza A, Ibarrondo O, et al. Validation of Random Forest Machine Learning Models to
486 Predict Dementia-Related Neuropsychiatric Symptoms in Real-World Data. *J Alzheimers Dis*
487 2020;77(2):855-64. doi:10.3233/JAD-200345
- 488 30. Iwendi C, Bashir AK, Peshkar A, et al. COVID-19 Patient Health Prediction Using Boosted Random
489 Forest Algorithm. *Front Public Health* 2020;8:357. doi:10.3389/fpubh.2020.00357
- 490 31. Munteanu CR, Gutiérrez-Asorey P, Blanes-Rodríguez M, et al. Prediction of Anti-Glioblastoma
491 Drug-Decorated Nanoparticle Delivery Systems Using Molecular Descriptors and Machine Learning.
492 *Int J Mol Sci* 2021;22(21):11519. doi: 10.3390/ijms222111519.
- 493 32. Nikolaidis A, Solon Heinsfeld A, Xu T, et al. Bagging improves reproducibility of functional
494 parcellation of the human brain. *Neuroimage* 2020;214:116678. doi:10.1016/j.neuroimage.2020.116678
- 495 33. Li S, Zeng Y, Chapman WC Jr, et al. Adaptive Boosting (AdaBoost)-based multiwavelength spatial
496 frequency domain imaging and characterization for ex vivo human colorectal tissue assessment. *J*
497 *Biophotonics* 2020;13(6):e201960241. doi:10.1002/jbio.201960241

- 498 34. Car Z, Baressi Šegota S, Anđelić N, et al. Modeling the Spread of COVID-19 Infection Using a
499 Multilayer Perceptron. *Comput Math Methods Med* 2020;2020:5714714. doi:10.1155/2020/5714714
- 500 35. Leithner D, Mayerhoefer ME, Martinez DF, et al. Non-Invasive Assessment of Breast Cancer
501 Molecular Subtypes with Multiparametric Magnetic Resonance Imaging Radiomics. *J Clin Med*
502 2020;9(6):1853. doi:10.3390/jcm9061853
- 503 36. Kang M, Shin S, Zhang G, et al. Mental Stress Classification Based on a Support Vector Machine
504 and Naive Bayes Using Electrocardiogram Signals. *Sensors (Basel)* 2021;21(23):7916. doi:
505 10.3390/s21237916.
- 506 37. Tewari P, Kashdan E, Walsh C, et al. Estimating the conditional probability of developing human
507 papilloma virus related oropharyngeal cancer by combining machine learning and inverse Bayesian
508 modelling. *PLoS Comput Biol.* 2021;17(8):e1009289. doi: 10.1371/journal.pcbi.1009289.
- 509 38. Varghese J, Fujarski M, Hahn T, et al. The Smart Device System for Movement Disorders:
510 Preliminary Evaluation of Diagnostic Accuracy in a Prospective Study. *Stud Health Technol Inform*
511 2020;270:889-93. doi:10.3233/SHTI200289
- 512 39. Satoh Y, Tamada D, Omiya Y, et al. Diagnostic Performance of the Support Vector Machine Model
513 for Breast Cancer on Ring-Shaped Dedicated Breast Positron Emission Tomography Images. *J Comput*
514 *Assist Tomogr* 2020;44(3):413-8. doi:10.1097/RCT.0000000000001020
- 515 40. Lown M, Brown M, Brown C, et al. Machine learning detection of Atrial Fibrillation using wearable
516 technology. *PLoS One* 2020;15(1):e0227401. doi:10.1371/journal.pone.0227401
- 517 41. Angelis I, Exarchos T. Hepatocellular Carcinoma Detection Using Machine Learning Techniques.
518 *Adv Exp Med Biol* 2021;1338:21-29. doi: 10.1007/978-3-030-78775-2_4.
- 519 42. Jagadev P, Giri LI. Human respiration monitoring using infrared thermography and artificial
520 intelligence. *Biomed Phys Eng Express* 2020;6(3):035007. doi: 10.1088/2057-1976/ab7a54.

- 521 43. Verma V, Vishwakarma RK, Verma A, et al. Time-to-Death approach in revealing Chronicity and
522 Severity of COVID-19 across the World. PLoS One 2020;15(5):e0233074.
523 doi:10.1371/journal.pone.0233074
- 524 44. Ghosal S, Sengupta S, Majumder M, et al. Linear Regression Analysis to predict the number of
525 deaths in India due to SARS-CoV-2 at 6 weeks from day 0 (100 cases - March 14th 2020). Diabetes
526 Metab Syndr 2020;14(4):311-5. doi:10.1016/j.dsx.2020.03.017
- 527 45. Ma X, Ng M, Xu S, et al. Development and validation of prognosis model of mortality risk in patients
528 with COVID-19. Epidemiol Infect 2020;148:e168. doi:10.1017/S0950268820001727
- 529 46. Tabaie A, Nemati S, Allen JW, et al. Assessing Contribution of Higher Order Clinical Risk Factors
530 to Prediction of Outcome in Aneurysmal Subarachnoid Hemorrhage Patients. AMIA Annu Symp Proc
531 2020;2019:848-56.
- 532 47. Lupei MI, Li D, Ingraham NE, et al. A 12-hospital prospective evaluation of a clinical decision
533 support prognostic algorithm based on logistic regression as a form of machine learning to facilitate
534 decision making for patients with suspected COVID-19. PLoS One 2022;17(1):e0262193. doi:
535 10.1371/journal.pone.0262193.
- 536 48. Deng Q. Dynamics and Development of the COVID-19 Epidemic in the United States: A
537 Compartmental Model Enhanced with Deep Learning Techniques. J Med Internet Res
538 2020;22(8):e21173. doi: 10.2196/21173
- 539 49. Jung SY, Jo H, Son H, et al. Real-World Implications of a Rapidly Responsive COVID-19 Spread
540 Model with Time-Dependent Parameters via Deep Learning: Model Development and Validation. J Med
541 Internet Res 2020;22(9):e19907. doi: 10.2196/19907
- 542 50. Angenent-Mari NM, Garruss AS, Soenksen LR, et al. A deep learning approach to programmable
543 RNA switches. Nat Commun 2020;11(1):5057. doi:10.1038/s41467-020-18677-1

- 544 51. Ribeiro AH, Ribeiro MH, Paixão GMM, et al. Automatic diagnosis of the 12-lead ECG using a deep
545 neural network. *Nat Commun* 2020;11(1):1760. doi:10.1038/s41467-020-15432-4
- 546 52. Banerjee A, Ray S, Vorselaars B, et al. Use of Machine Learning and Artificial Intelligence to predict
547 SARS-CoV-2 infection from Full Blood Counts in a population. *Int Immunopharmacol* 2020;86:106705.
548 doi:10.1016/j.intimp.2020.106705
- 549 53. Takao H, Amemiya S, Kato S, et al. Deep-learning single-shot detector for automatic detection of
550 brain metastases with the combined use of contrast-enhanced and non-enhanced computed tomography
551 images. *Eur J Radiol* 2021;144:110015. doi: 10.1016/j.ejrad.2021.110015.
- 552 54. Pokaprakarn T, Kitzmiller RR, Moorman R, et al. Sequence to Sequence ECG Cardiac Rhythm
553 Classification using Convolutional Recurrent Neural Networks. *IEEE J Biomed Health Inform* 2021;
554 doi: 10.1109/JBHI.2021.3098662.
- 555 55. Winkler JK, Sies K, Fink C, et al. Melanoma recognition by a deep learning convolutional neural
556 network-Performance in different melanoma subtypes and localisations. *Eur J Cancer* 2020;127:21-9.
557 doi:10.1016/j.ejca.2019.11.020
- 558 56. Sedik A, Iliyasu AM, Abd El-Rahiem B, et al. Deploying Machine and Deep Learning Models for
559 Efficient Data-Augmented Detection of COVID-19 Infections. *Viruses* 2020;12(7):769. doi:
560 10.3390/v12070769
- 561 57. Ko H, Chung H, Kang WS, et al. COVID-19 Pneumonia Diagnosis Using a Simple 2D Deep
562 Learning Framework with a Single Chest CT Image: Model Development and Validation. *J Med Internet
563 Res* 2020;22(6):e19569. doi:10.2196/19569
- 564 58. Oberai A, Varghese B, Cen S, et al. Deep learning based classification of solid lipid-poor contrast
565 enhancing renal masses using contrast enhanced CT. *Br J Radiol* 2020;93(1111):20200002.
566 doi:10.1259/bjr.20200002

- 567 59. Wetstein SC, Onken AM, Luffman C, et al. Deep learning assessment of breast terminal duct lobular
568 unit involution: Towards automated prediction of breast cancer risk. *PLoS One* 2020;15(4):e0231653.
569 doi:10.1371/journal.pone.0231653
- 570 60. Arshad M, Khan MA, Tariq U, et al. A Computer-Aided Diagnosis System Using Deep Learning for
571 Multiclass Skin Lesion Classification. *Comput Intell Neurosci* 2021;2021:9619079.
572 doi:10.1155/2021/9619079
- 573 61. Mammone N, Ieracitano C, Morabito FC. A deep CNN approach to decode motor preparation of
574 upper limbs from time-frequency maps of EEG signals at source level. *Neural Netw* 2020;124:357-72.
575 doi:10.1016/j.neunet.2020.01.027
- 576 62. Muramatsu C. Diagnosis of Glaucoma on Retinal Fundus Images Using Deep Learning: Detection
577 of Nerve Fiber Layer Defect and Optic Disc Analysis. *Adv Exp Med Biol* 2020;1213:121-32.
578 doi:10.1007/978-3-030-33128-3_8
- 579 63. Hatanaka Y. Retinopathy Analysis Based on Deep Convolution Neural Network. *Adv Exp Med Biol*
580 2020;1213:107-20. doi: 10.1007/978-3-030-33128-3_7
- 581 64. Ganitidis T, Athanasiou M, Dalakleidi K, et al. Stratification of carotid atheromatous plaque using
582 interpretable deep learning methods on B-mode ultrasound images. *Annu Int Conf IEEE Eng Med Biol*
583 *Soc* 2021;2021:3902-3905. doi: 10.1109/EMBC46164.2021.9630402.
- 584 65. Kaur P, Harnal S, Tiwari R, et al. A Hybrid Convolutional Neural Network Model for Diagnosis of
585 COVID-19 Using Chest X-ray Images. *Int J Environ Res Public Health* 2021;18(22):12191. doi:
586 10.3390/ijerph182212191.
- 587 66. Golestani N, Moghaddam M. Human activity recognition using magnetic induction-based motion
588 signals and deep recurrent neural networks. *Nat Commun* 2020;11(1):1551. doi: 10.1038/s41467-020-
589 15086-2

- 590 67. Ljubic B, Abdel Hai A, Stanojevic M, et al. Predicting Complications of Diabetes Mellitus Using
591 Advanced Machine Learning Algorithms. *JAMIA* 2020;27(9):1343-51. doi:10.1093/jamia/ocaa120
- 592 68. Lobo B, Abdel-Rahman E, Brown D, et al. A recurrent neural network approach to predicting
593 hemoglobin trajectories in patients with End-Stage Renal Disease. *Artif Intell Med* 2020;104:101823.
594 doi:10.1016/j.artmed.2020.101823
- 595 69. Li R, Wu Q, Liu J, et al. Monitoring Depth of Anesthesia Based on Hybrid Features and Recurrent
596 Neural Network. *Front Neurosci* 2020;14:26. doi:10.3389/fnins.2020.00026
- 597 70. Ljubic B, Roychoudhury S, Cao XH, et al. Influence of medical domain knowledge on deep learning
598 for Alzheimer's disease prediction. *Comput Methods Programs Biomed* 2020;197:105765.
599 doi:10.1016/j.cmpb.2020.105765
- 600 71. Choe S, Park E, Shin W, et al. Short-Term Event Prediction in the Operating Room (STEP-OP) of
601 Five-Minute Intraoperative Hypotension Using Hybrid Deep Learning: Retrospective Observational
602 Study and Model Development. *JMIR Med Inform* 2021;9(9):e31311. doi: 10.2196/31311.
- 603 72. Kim K, Yang H, Yi J, Son HE, et al. Real-Time Clinical Decision Support Based on Recurrent Neural
604 Networks for In-Hospital Acute Kidney Injury: External Validation and Model Interpretation. *J Med*
605 *Internet Res* 2021;23(4):e24120. doi: 10.2196/24120.
- 606 73. Ioannou GN, Tang W, Beste LA, et al. Assessment of a Deep Learning Model to Predict
607 Hepatocellular Carcinoma in Patients with Hepatitis C Cirrhosis. *JAMA Netw Open*
608 2020;3(9):e2015626. doi: 10.1001/jamanetworkopen.2020.15626.
- 609 74. <https://www.hcup-us.ahrq.gov/databases.jsp>.
- 610 75. Antun V, Renka F, Poon C, et al. On instability of deep learning in image reconstruction and the
611 potential costs of AI. *Proceedings of the National Academy of Sciences of the United States of America*
612 2020;117(48):30088-95

- 613 76. Stojkovic I, Jelisavcic V, Milutinovic V, et al. Distance Based Modeling of Interactions in Structured
614 Regression. In IJCAI 2016;2032-8.
- 615 77. Stojkovic I, Jelisavcic V, Milutinovic V, et al. Fast Sparse Gaussian Markov Random Fields Learning
616 Based on Cholesky Factorization. In IJCAI 2017;2758-64.