


1 Applying Multimodal Data Fusion based on Deep Learning Methods for the 2 Diagnoses of Neglected Tropical Diseases: A Systematic Review

3 *Yohannes Minyilu*^{1*}, *Mohammed Abebe*¹, *Million Meshesha*²

4 ¹ Faculty of Computing and Software Engineering, Institute of Technology, Arba Minch University, Arba Minch, Ethiopia, ² Department of
5 Information Science, School of Information Science, Addis Ababa University, Addis Ababa, Ethiopia

6 *yohannes.minyilu@gmail.com

7  All the authors contributed equally to this work

8 Abstract

9 *Neglected tropical diseases (NTDs) are the most prevalent diseases worldwide affecting one-*
10 *tenth of the world population. Although there are multiple approaches to diagnosing these*
11 *diseases, using skin manifestations and lesions caused as a result of these diseases along with*
12 *other medical records is the preferred method. This fact triggers the need to explore and*
13 *implement a deep learning-based diagnostic model using multimodal data fusion (MMDF)*
14 *techniques to enhance the diagnostic process. This paper, thus, endeavored to present a*
15 *thorough systematic review of studies regarding the implementation of MMDF techniques for the*
16 *diagnoses of skin-related NTDs. To achieve its objective, the study used the PRISMA method*
17 *based on predefined questions and collected 427 articles from seven major and reputed sources*
18 *and critically appraised each article. Since no previous studies were found regarding the*
19 *implementation of MMDF for the diagnoses of skin related NTDs, similar studies using MMDF*
20 *for the diagnoses of other skin diseases, such as skin cancer, were collected and analyzed in this*
21 *review to extract information about the implementation of these methods. In doing so, various*
22 *studies are analyzed using six different parameters including research approaches, disease*
23 *selected for diagnosis, dataset, algorithms, performance achievements and future directions.*
24 *Accordingly, although all the studies used diverse research methods and datasets based on their*
25 *problem, deep learning-based convolutional neural networks (CNN) algorithms are found to be*
26 *the most frequently used and best performing models in all studies reviewed.*

27 **Key words:** *Deep Learning, Disease Diagnosis, Multimodal Data Fusion, Skin NTDs,*

28 **1. Introduction**

29 Being the largest organ in the human body, the skin can serve as an indicator of some illnesses
30 arising from different causes such as cancer, internal organ failure, and neglected tropical
31 diseases. NTDs are the most prevalent diseases globally affecting more than one billion people
32 worldwide (i.e., more than ten percent of the world's population), particularly, in the tropical
33 areas of the world among the poorest, most vulnerable and outcast groups and still have
34 devastating impacts on people's physical, mental, and social well-being [1][2][3][4]. However,
35 these diseases can be diagnosed using skin related symptoms since majority of the NTDs have
36 primary skin indicators or associated clinical features where 18 of the 20 NTDs (recognized by
37 the World Health Organization (WHO)) having skin related symptoms [5]. Hence, the utilization
38 of DL-based diagnostic systems for the diagnoses and recognition of skin related NTDs will be a
39 great achievement in overcoming the NTDs. This study endeavored to present a thorough
40 systematic review of studies regarding the implementation of MMDF techniques for the
41 diagnoses of skin related NTDs. Since no previous studies implemented MMDF techniques for
42 the diagnoses of skin NTDs, related studies conducted for the diagnoses of skin diseases other
43 than NTDs using MMDF and DL methods were deeply appraised by this review. These studies
44 confirmed that the utilization of MMDF techniques outperforms the traditional diagnostic
45 models that implemented DL methods without MMDF [6][7][8][9]. It is in view of these facts
46 the study is -motivated to conduct this systematic literature review and a thorough appraisal of
47 previous studies using the PRISMA method of systematic review based on the following guiding
48 questions:

- 49 • What DL methods or approaches were utilized for the diagnoses of the skin disease(s)?
- 50 • Which data fusion methods were used for the skin disease diagnosis tasks?
- 51 • What types of medical data were integrated to demonstrate MMDF method for the
52 diagnoses of the skin diseases?
- 53 • Which algorithms were used and how does each algorithm perform in the DL-based
54 MMDF skin disease diagnostic model or system?

55 **2. The Need for Intelligent Diagnostic Systems**

56 In recent times, due to the high desire to enhance the diagnostic processes in the healthcare
57 sectors, the utilization of automated and intelligent diagnostic systems is getting greater attention
58 for the diagnosis of various diseases. In this regard, intelligent diagnostic systems built based on
59 machine learning (ML) and deep learning (DL) methods are the most researched and deployed
60 approaches in the healthcare sector to support diagnostic decision making. On the other hand, in
61 the real-world clinical settings, efficient disease diagnostic processes are basically carried out by
62 using different clinical data that are taken from different sources and different formats or
63 modalities including textual patient information and medical clinical images such as X-ray,
64 dermoscopic images or even patient skin images. The integrative utilization of the diverse
65 modalities of medical data can be used to enhance the diagnostic processes, thereby enhancing
66 the quality of healthcare services, by using the ML and DL methods. In ML, this process of
67 integrating multiple modalities of data (possibly taken from different sources) is technically
68 called multimodal data fusion [10][11]. Multimodal data techniques are playing vital roles in
69 developing intelligent disease diagnostic systems for different diseases such as in dermatology
70 [12]. In this regard, MMDF techniques are advancing diagnostic accuracy, where these methods
71 outperform other baseline methods, as presented in [13].

72 **2.1. Deep Learning and Diagnoses of NTDs**

73 The current diagnostic approaches used for NTDs are mainly based on clinical procedures, such
74 as patient observation and laboratory examinations based on limited resources in most affected
75 areas. Currently, however, there are efforts towards the utilization of intelligent diagnostic tools
76 using ML and DL approaches. Since most of the NTDs are curly being diagnosed using skin
77 manifestation, the utilization of DL-based approaches for the diagnoses of these diseases would
78 be a great potential to support and enhance the diagnostic processes. In this regard, different
79 studies were previously conducted to diagnose various NTDs. Beesetty et al. [14], conducted a
80 study towards leprosy skin lesion detection employing a Siamese (Siamese NN)-based few-shot
81 learning (FSL) model for a small clinical dataset and claimed a higher (91.25%) diagnostic
82 accuracy. On the other hand, Ali et al. [15] used ML methods for early prediction of
83 Schistosomiasis and concluded with the CatBoost model showing the best performance with the

84 highest accuracy of 87.1%. An optimized diagnostic approach was also proposed for NTDs by
85 selecting three diseases and developing a model using SVM and the black hole algorithm (BHO)
86 achieving an accuracy of 96% [16]. Another study reviewed by the current study demonstrated a
87 DL-based diagnostic model for NTDs using skin images only and achieved 70% accuracy [17].
88 All the aforementioned studies utilized ML and DL methods for the diagnoses of NTDs and
89 achieved remarkable results in terms of accuracy. However, no previous studies were found that
90 utilized DL-based methods using MMDF techniques which will help to achieve higher
91 diagnostic accuracies, as experimented in other studies for non-NTD skin diseases which require
92 further research.

93 **2.2. Data Fusion Approaches**

94 Data or information fusion represents the usage of data or information from different sources in
95 different formats or modalities for interpretation in all tasks that require any type of parameter
96 estimation or prediction using data or information [18]. There are different fusion techniques to
97 combine and aggregate multimodal data which include feature-level fusion, decision-level or late
98 fusion, hybrid multimodal fusion, model-level fusion, rule-based fusion, classification-based
99 fusion and estimation-based fusion [19].

100 **2.2.1. Feature Fusion**

101 Feature fusion is a data integration technique used to aggregate multiple feature sets extracted
102 from multiple input data to generate a single feature set [19]. In image processing problems, it
103 refers to the fusion of feature vectors of training images extracted from shared weight network
104 layer and feature vectors composed of other numerical data [20]. It helps to learn image features
105 fully for the description of their rich internal information [21]. Various studies are found and appraised
106 that use feature fusion techniques to develop diagnostic models for the diagnoses of skin diseases, as
107 summarized in Table 1.

108 **2.2.2. Model Fusion**

109 Model fusion, also known as late fusion, represents a fusion approach that combines different
110 models. The study done by AlDahoul et al. [22], combines two deep neural networks including

111 binary normal/attack classifier and multi-attack classifier to train a deep neural network (DNN)
112 for network anomaly detection. As mentioned in [19], the model fusion technique uses the
113 connection between experimental data under different modalities.

114 **2.2.3. Image Fusion**

115 Image fusion combines different images and generates informative images by integrating images
116 obtained from different sources [23]. A previous study [24], suggested that aggregating medical
117 images helps to enhance diagnostic accuracy. This claim was demonstrated by fusing clinical
118 images and dermoscopic images using the deep convolutional neural networks (DCNN) methods
119 and achieved an overall accuracy of 81.3%. While the clinical images are clinically captured
120 photographs [25], dermoscopic images represent images taken by dermatologists using
121 dermoscopy [26].

122 **2.2.4. Multimodal Data Fusion**

123 Multimodal data represents the different formats or modalities of data such as text, image, video,
124 and audio. A multimodal data fusion approach is used for combining particular modalities to
125 derive multimodal representation [10][11][19][27]. This approach has multiple applications for
126 healthcare systems as it allows the combination of different modalities of data, such as textual
127 medical history of patients, clinical images of patients (such as skin images of patients) to form a
128 single multimodal dataset that can be used to train diagnostic models using ML and DL methods.
129 In this regard, various studies implemented and demonstrated MMDF for the diagnoses of
130 different skin diseases as summarized in Table 2.

131 **3. Materials and Methods**

132 To conduct and report this systematic review, we follow as a basis, the steps suggested by the
133 Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) model
134 [28][29] as described below.

135 **3.1. Search Strategy**

136 Research articles were searched from all the major indexing databases and search engines, such
137 as Google Scholar, PubMed, IEEE Explore, and ScienceDirect. In addition, proper search
138 keywords were prepared and used while searching for the articles, where the search keywords
139 have the appropriate level of relationships with the topics and contents of the articles. A set of
140 searching keywords have been used to deeply search and filter the articles. In this regard,
141 Boolean operators “AND” and “OR” were mainly used. The “AND” operator was used to search
142 for articles in a specific research area to narrow down the search results; “multimodal medical
143 data” AND “data fusion” to search for articles containing both phrases. On the other hand, the
144 “OR” Boolean operator was used to search for articles from wider perspectives as this operator
145 broadens the search results such as “machine learning” OR “deep learning”.

146 Using the specified methods and operators, multiple search keywords were initially prepared and
147 used to find a sufficient amount of relevant articles. Although a lot of search keywords were used
148 while searching for the articles, some of the keywords include ["Neglected Tropical Diseases"
149 OR "NTDs" AND "Diagnosis" OR "diagnostic model" AND "Deep Learning" OR "DL" OR
150 "Convolutional Neural Network" OR "CNN" OR "Deep Neural Network" OR "DNN" OR
151 "Recurrent Neural Network" OR "RNN"], ["Neglected Tropical Diseases" OR "NTDs" AND
152 "Diagnosis" OR "diagnostic model" AND "Deep Learning" OR "DL" or "Convolutional Neural
153 Network" OR "CNN" OR "Deep Neural Network" OR "DNN" OR "Recurrent Neural Network"
154 OR "RNN" AND "Data Fusion" OR "Multimodal medical Data" OR "Multimodal Data
155 Fusion"], [(((deep learning) AND ((diagnostic model) OR (diagnostic system) OR (diagnostic
156 tool)) AND (skin diseases) AND (skin images)) AND (medical record)) AND (data fusion)].

157 **3.2. Eligibility Criteria**

158 Out of the total 427 selected articles, not all articles are critically relevant for the review
159 concerning the integration of multimodal data fusion techniques based on DL methods for the
160 diagnosis of skin related NTDs. Hence, a set of inclusion and/or exclusion criteria are applied, as
161 shown below.

- 162 ❖ Articles must use and demonstrate DL methods for the diagnosis of skin NTDs, at least
163 skin diseases if not implemented for skin NTDs, with proper evaluation of the methods
164 used.
- 165 ❖ Articles must use and demonstrate proper utilization of multimodal data fusion
166 techniques for the diagnosis of skin NTDs, at least skin diseases since there are no
167 previous studies that use multimodal data fusion techniques for the diagnosis of skin
168 NTDs so far.
- 169 ❖ Articles must incorporate precise presentation or discussion and evaluation of all the
170 methods and techniques used in that particular article.
- 171 ❖ An Article that used DL methods for the diagnosis of skin related diseases other than
172 NTDs is selected if that particular article uses new or emerging DL methods and presents
173 a proper analysis of the methods and techniques used for the diagnosis of that particular
174 skin disease(s). However, articles that use the popular and previously used DL methods
175 for the diagnosis of diseases other than skin diseases have fewer chances to be selected.
- 176 ❖ The article should have an appropriate level of similarity and relationships in its topics
177 and contents with the searching keywords used to deeply search and filter the articles.
- 178 ❖ Articles that do not utilize DL and data fusion techniques are excluded from analysis.
- 179 ❖ Articles published in languages other than English are excluded from analysis.
- 180 ❖ Finally, articles published prior to the year 2014 are also excluded.

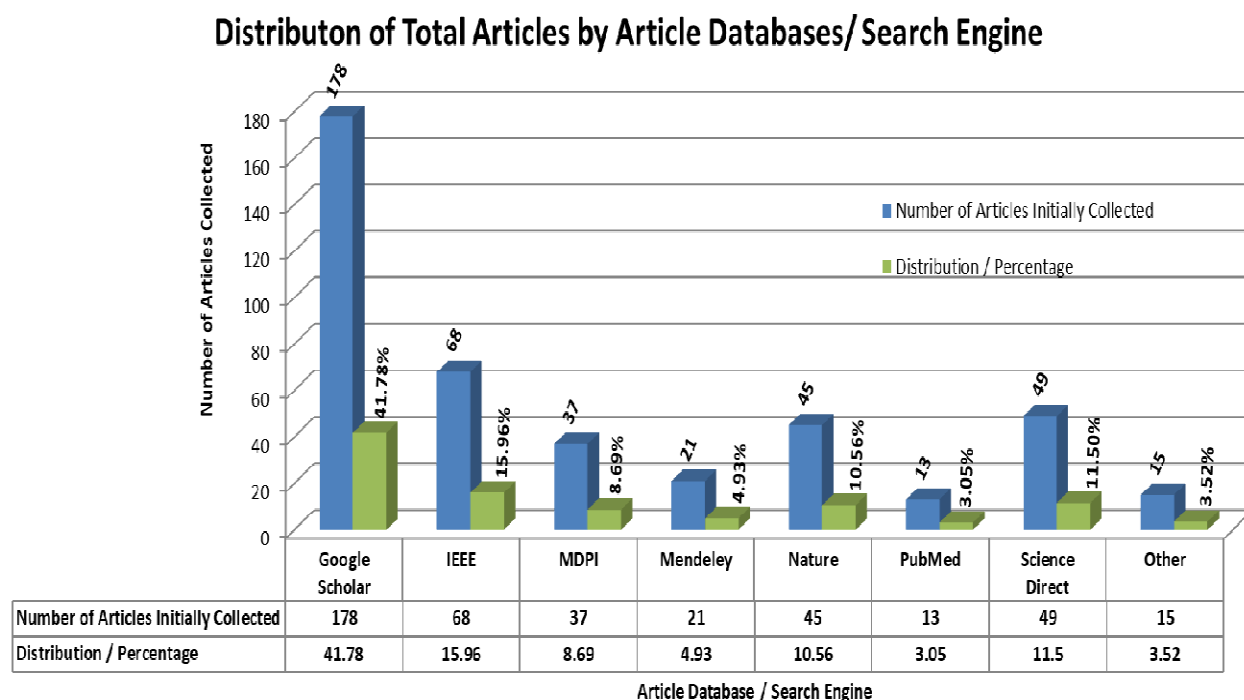
181 **3.3. Article Search**

182 Different searching methods such as ‘basic search’ and ‘advanced search’ methods were used on
183 multiple article sources. First, the ordinary or basic searching method was used where general
184 titles and the proposed keywords were entered in the regular ‘search box’ of each of the
185 databases and searched. Secondly, the ‘advanced search’ option was used which allows to
186 specify subject areas, related topics, publication dates and other relevant options which helps to
187 obtain articles that are relevant to the topic by narrowing down the search results.

188 Using both of the search methods and search keywords, a thorough and rigorous searching was
189 conducted on multiple search engines, journals, databases and libraries to find relevant articles.
190 The sources include Google Scholar, IEEE Explore, MDPI, Mendeley, Nature, PubMed,

191 ScienceDirect, AJOL, IDP, NCBI, PLOS, Springer, and Tropical Medicine and Health. Finally,
 192 by specifying article publication dates and applying the searching methods on the different
 193 databases, 427 articles that were published between the year 2014 and 2024 were collected and
 194 prepared for screening. Each database was used independently to search articles. In this regard,
 195 Google Scholar was primarily used and it allowed us to collect 178 articles from the
 196 aforementioned different sources. Furthermore, extensive searches were done in search of
 197 articles that implement multimodal data fusion for the diagnosis of skin NTDs. However, no
 198 relevant articles were found related to this particular area.

199 On the other hand, previously searched sources such as academic web portals, academic libraries
 200 and research sites were also used as there were relevant documents from these sources. Hence,
 201 16 articles were collected from such sources.



202

203 **Fig 1: Total collected articles and their distribution by article databases/ search engines**

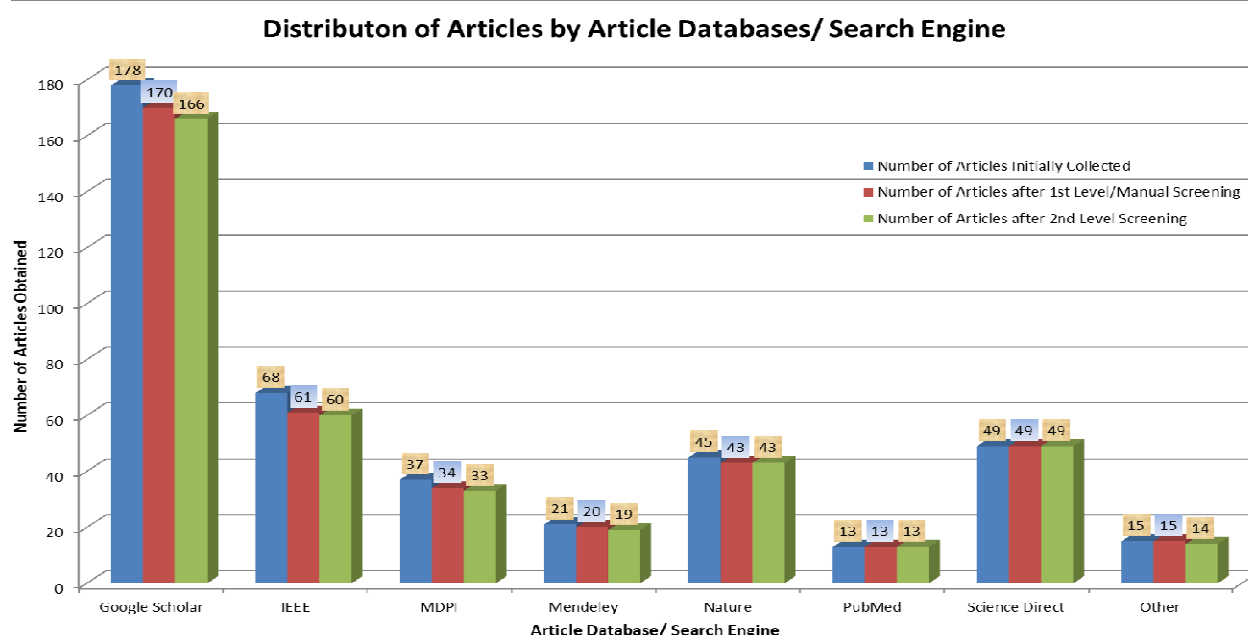
204 **3.4. Relevant Article Selection**

205 To select relevant articles, an extensive searching method was used using wider options of
 206 searching keywords. The entire process of article selection for this review was conducted based

207 on the PRISMA method as it was an evidence-based minimum set of items for reporting for
208 systematic reviews and meta-analyses [29][28]. This was done to provide insights for the recent
209 and future research regarding the utilization of DL methods for NTDs diagnosis, the integration
210 of data fusion techniques if they have been used for NTDs and finally to assess and present
211 feedback so as to enhance the performances of such DL-based models.

212 A series of screening operations were implemented on the collected articles in order to identify
213 the most relevant set of articles for this review. In this case, the first level screening was
214 conducted manually on a total of 427 files using file names and titles of the article. This task
215 allowed us to check if there were duplicates as there were similar files downloaded from multiple
216 sources and it was performed manually by opening and checking files. In this process, 397 items
217 were selected out of the total 427 articles in three phases as summarized in Figure 2 below.

218 Then, the next levels of screenings were performed using software tools such as ‘EndNote’ and
219 ‘Rayyan’. As a reference management tool, EndNote was used to create a library containing the
220 collected articles and for manipulation and data processing to check duplicate files in the library.
221 It automatically removed 25 articles as there were duplicate files from different folders followed
222 by an automatic duplicate detection where one duplicate article was identified by EndNote and
223 removed leading to a library containing 371 articles.

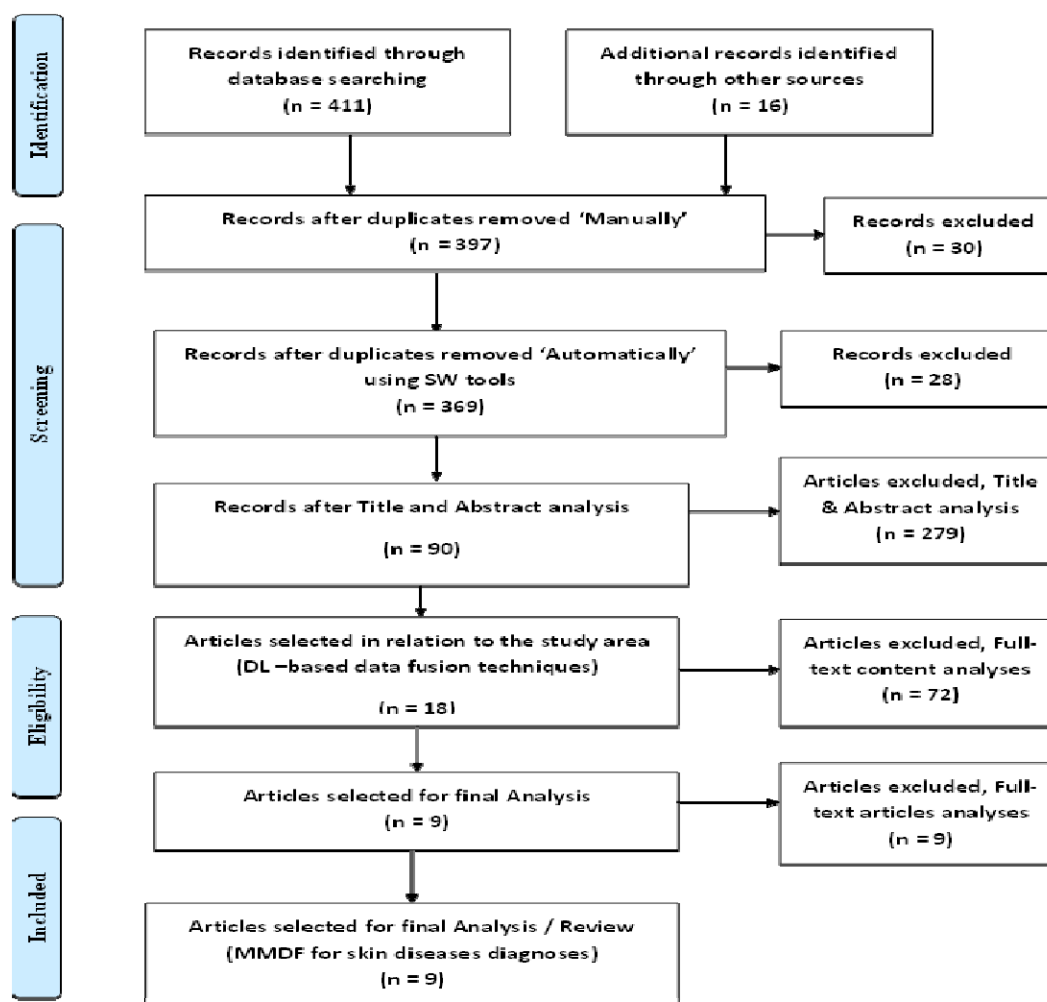


224

225

Fig 2: Distribution of articles after first and second level screening

226 The next task involved screening the articles using a higher-level screening software tool based
227 on title, author names and abstracts. For this purpose, Rayyan, a free online software tool [30]
228 was employed which is mainly used to speed up the literature screening process in systematic
229 reviews. This online tool uses the article library exported from EndNote and it was first used to
230 check duplication. Through this process, 4 duplicated articles were detected in the library and
231 two of them were removed where 369 articles were finally identified using Rayyan for the final
232 screening process. Next, using this online software tool, 90 articles that have a relationship with
233 the current topic of the study were selected based on title and abstract analysis. Further screening
234 was required to identify articles in relation to the study area and 18 articles were identified out of
235 the 90 related articles. Finally, 9 articles were selected for the final analysis. The overall article
236 selection procedure is outlined using the PRISMA flow chart as depicted in Figure 3 below.

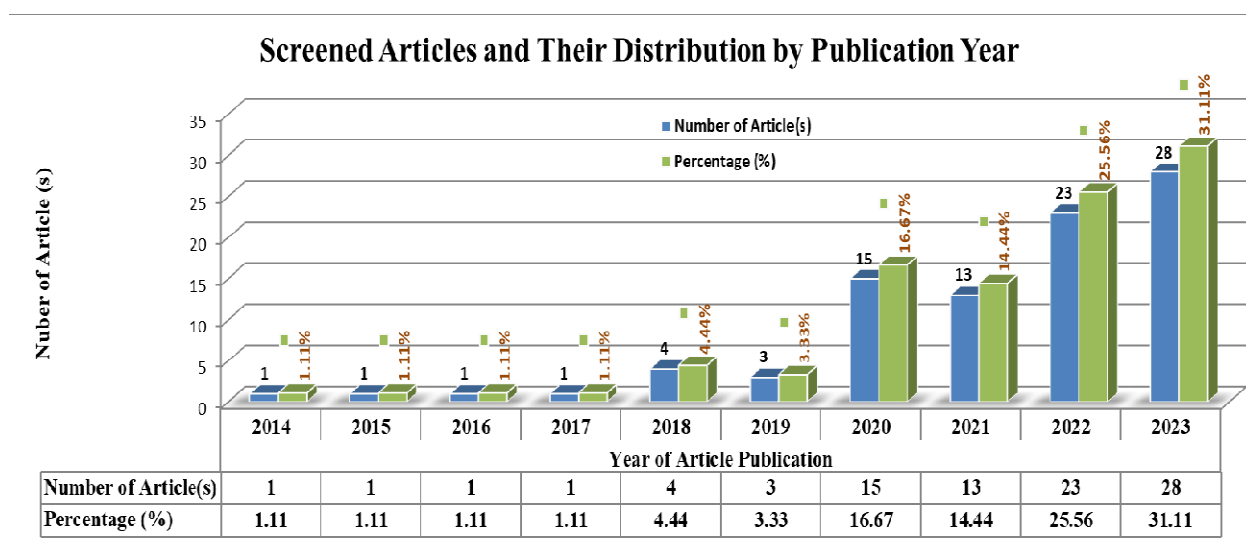


237

238 **Fig 3: Article selection for systematic literature review following PRISMA flow chart**

239 4. Results

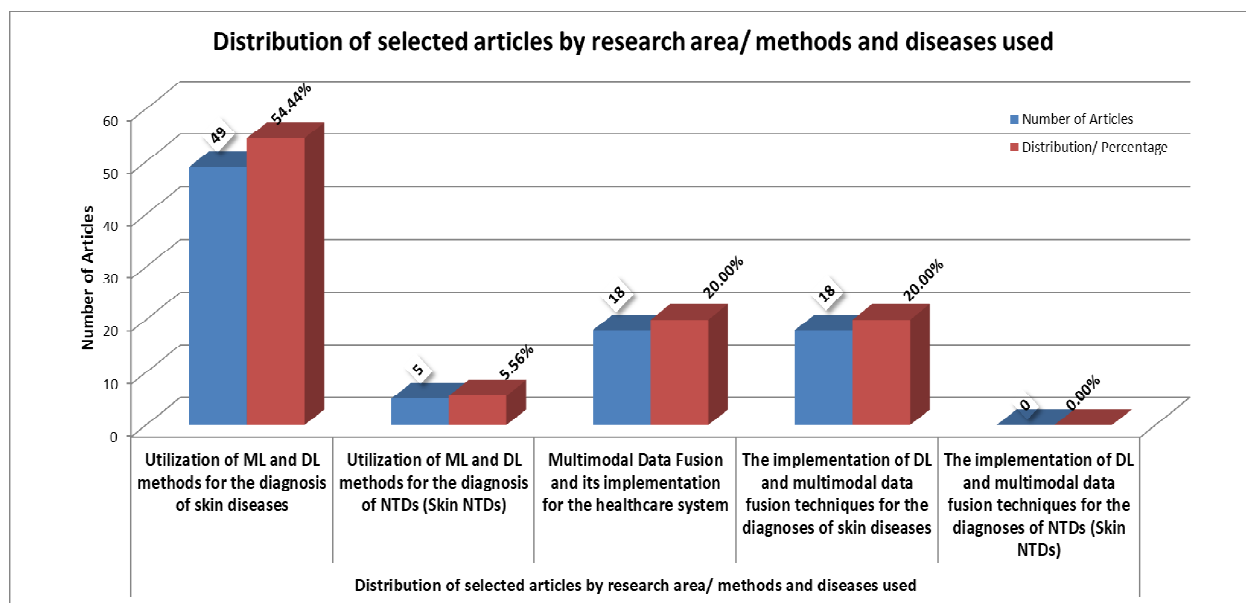
240 After conducting three levels of screening, 90 articles that have a direct relationship with the
 241 current systematic review have been selected for further screening based on full-text reading and
 242 analyses. The selected articles and their respective publication year along with the distribution of
 243 the publications years have been shown in Figure 4 below. As shown in Figure 4, the articles
 244 used for this systematic review included studies that have been published recently, where the
 245 majority of the studies representing 31% are articles published in 2023, 25% were published in
 246 2022, 16% were published in 2021, 14% were published in 2020, and the remaining 14% were
 247 articles published from 2014 – 2019.



248

249 **Fig 4: Distribution of articles after the third level screening by publication year**

250 Finally, the 90 articles were further analyzed by categorizing them into four different groups, (i)
 251 articles that utilized DL methods for the diagnosis of skin diseases, (ii) articles that implement
 252 ML & DL techniques for the diagnosis of NTDs, (iii) articles about the implementation of
 253 multimodal data fusion techniques for medical data fusion, and (iv), articles that implement
 254 multimodal data fusion based on DL-based methods for the diagnosis of skin diseases as shown
 255 in Figure 5 below.



256

257 **Fig 5: Distribution of articles by research area/methods and selected diseases after the third level screening**

258 As portrayed in Figure 5 above, 54.44% of articles utilized ML and DL methods for the
 259 diagnosis of skin diseases in general, 20% deal with multimodal data fusion techniques for
 260 healthcare systems and 20% implementation of DL-based multimodal data fusion methods for
 261 the diagnosis of skin diseases. On the other hand, 5.56% of the articles utilized ML and DL
 262 methods for the diagnoses of NTDs in general have been identified and analyzed. However, no
 263 article has been found that deals with the implementation of DL-based MMDF methods for the
 264 diagnosis of NTDs which has led to the analyses of previous studies that used this approach for
 265 the diagnoses of different skin diseases other than the NTDs. By conducting the fourth level
 266 screening, 18 articles that utilize different fusion techniques for the diagnosis of various skin
 267 diseases have been identified.

268 4.1. Article Analysis

269 The final screening has resulted in the separation of 7 of the 18 articles due to the fusion
 270 techniques they utilize for the diagnosis of skin diseases. The fusion techniques presented in
 271 those 7 studies are feature fusion (5 studies), image fusion (1 study) and model fusion (1 review
 272 study) as presented in Table 1 below. Table 1 presented the analysis of three different types of
 273 fusion other than MMDF using five different parameters as shown in the table below.

274 **Table 1: Review of the future fusion and related techniques for skin disease diagnoses**

Ref	Pub. Yr.	Study Method / Approach Used	Disease(s) Selected	Dataset(s) Used	Algorithm(s) Used	Performance Results Achieved
[31]	2019	Transfer Learning and multi-layer feature fusion network	Skin Lesion	HAM1000 0 dataset	CNN	high recognition (ROC-AUC 96.51)
[24]	2021	Image fusion (clinical & dermoscopic) : multi-labeled deep feature extractor and clinically constrained classifier chain (CC)	Skin Cancer (Melanoma)	publicly available 7-point checklist dataset	DCNN, CC, PCA	Reported 81.3% accuracy
[6]	2022	Multiclass skin lesion classification using feature fusion & extreme learning machine (ELM)	Skin Disease (Skin Lesion)	HAM1000 0 and ISIC2018	SVM, fine KNN, DT, NB, ensemble tree (EBT), & single hidden layer ELM	Registered best accuracy of 94.36 percent
[32]	2022	Apply features fusion on manual and automatic feature extraction	Skin Cancer	DermIS dataset	CNN, LSTM, LBP, LBP, Inception V3	Achieved maximum accuracy of 99.4%
[33]	2023	Dual-branch (feature fusion) network using DCNN and Transformer branches for local and global feature extraction	Skin Disease (Skin Lesion)	Used a private dataset XJUSL	DCNN	Reducing the number of parameters by 11.17 M improved classification accuracy by 1.08%
[34]	2023	Feature fusion : fast-bounding box (FBB), Hybrid Feature Extractor (HFE), and the CNN VGG19 based CNN	Skin Cancer (Melanoma)	ISIC 2017 and the Academic torrents dataset	CNN	Registered 99.85% accuracy

276 On the other hand, 2 articles presented a review of the multimodal data fusion techniques for the
277 diagnoses of skin diseases other than NTDs. Although the 2 articles [12][13], didn't implement
278 MMDF techniques for a specific skin disease diagnosis using their datasets of preferences, they
279 presented theoretical analyses. All in all, 9 articles are used for the final analysis of this review.

280 After conducting the final screening procedures, 9 articles have been selected for the final
281 analysis of this systematic review as presented in Table 2 below. The 9 articles selected utilized
282 DL-based methods based on MMDF techniques for the diagnoses of different skin diseases other
283 than NTDs. The 9 studies are selected for the final analysis of this review since there are no
284 similar studies found for the diagnosis of skin related NTDs based on MMDF. Since skin related
285 NTDs are being diagnosed using skin photos or images, patient records and related information,
286 these studies are selected and reviewed to analyze the different techniques utilized by those
287 studies. The final analysis is conducted on the 9 articles using 5 different analysis criteria (the
288 methods used, diseases selected for diagnosis, dataset used, algorithms used and corresponding
289 performance achievements) to identify research gaps as presented in Table 2 below.

290 **5. Discussion**

291 The primary goal of this systematic review was to collect and analyze research studies that are
292 pertinent to the area of DL-based models that use multimodal data fusion techniques for the
293 diagnosis of skin related NTDs. Since no studies were found in the specified area, similar or
294 related studies that implement MMDF techniques based on DL for the diagnoses of skin diseases
295 were collected and analyzed to extract pertinent information. In doing so, 9 articles (indexed in
296 Scopus and Web of Science) about data fusion techniques, particularly MMDF techniques for the
297 diagnosis of skin diseases have been examined and analyzed. Table 2 below presents the
298 summary of the studies that implemented the different MMDF techniques based on DL for the
299 diagnoses of different skin diseases.

Table 2: Review of the DL-based multimodal data fusion techniques for the diagnosis of skin diseases

Ref.	Pub. Yr.	Study Method / Approach Used	Disease(s) Selected	Dataset(s) Used	Algorithm(s) Used	Performance / Accuracy Results Achieved	Recommendations Forwarded
[35]	2018	Combining multiple imaging modalities (dermatoscopic & macroscopic) with patient metadata	5 cases (such as melanoma)	New DS composed of 2917 cases, from five classes (five selected diseases)	CNN, Random Forrest classifier, ResNet-50 arch., ILSVRC 2015	binary melanoma detection (AUC 0.866 vs 0.784) and in multiclass classification (mAP 0.729 vs 0.598)	Integrating more benign non-excised skin and further stratify them based on suspicion
[36]	2020	Multiplication-based DF, using the metadata	No specific disease selected	ISIC'2018 and three types of metadata	CNN, the color constancy algorithm	outperforms traditional baseline approaches (p-values are smaller than 0.05)	Exploring effects of more types of metadata for more skin diseases
[7]	2021	Combining images and metadata features: the (MetaBlock)	Skin Cancer	ISIC 2019 and PAD-UFES-20	CNN: using 5 pre-rained models	Performs better than the other combination approaches in 6 out of 10 scenarios.	N/A
[8]	2021	Performance analysis of classifiers, and a naive combination of patient data and an image classifier	Skin Cancer	Collected 431 WSIs from two different laboratories with patient information	CNN	CNN: AUROC of 92.30% $\pm 0.23\%$ & balanced accuracy of 83.17% $\pm 0.38\%$, naive strategy: accuracy to 86.72% $\pm 0.36\%$.	Results achieved need to be confirmed systematically in larger studies with diverse data sets.
[37]	2022	a DNN with two encoders and application of a multimodal fusion module with intra-modality self-attention and inter-modality cross-attention	Skin Cancer	PAD-UPES-20	CNN: CNN models (ResNet-50)	ACC (0.768 \pm 0.022) and BACC (0.775 \pm 0.022) & claimed outperforming other metadata fusion methods (MetaNet (P = 0.035) and MetaBlock (P = 0.028))	Integrating the model into smartphone as a potential and handy tool to screen for skin disease and skin cancer

medRxiv preprint doi: <https://doi.org/10.1101/2024.01.07.24300957>; this version posted January 16, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Table 2: Review of the DL-based multimodal data fusion techniques for the diagnosis of skin diseases

[38]	2022	Multimodal Transformer: Vision Transformer (ViT) model, Soft Label Encoder (SLE), and a Mutual Attention (MA) block	Skin disease	A private DS collected (760 images) & benchmark DS of the ISIC 2018	CNN: CNN models (ResNet101, Densenet121) and ViT models	Private DS (accuracy: 0.816, which is better than other popular networks) & On ISIC 2018 DS (accuracy: 0.9381 and an AUC of 0.99)	To focus on image feature extraction and also, the data collection and data cleaning of skin diseases
[39]	2022	Medical image analysis: Preprocessing, feature extraction, and classification/diagnosis, & the hold-out technique to split the dataset	7 skin diseases	HAM10000 dataset	CNN: 6 CNN pre-trained models, Hyper-Parameter Optimization (HPO) algorithms	Av. acc, sensitivity, specificity, precision, & disc similarity coefficient (DSC) of around 99.94%, 91.48%, 98.82%, 97.01%, and 94.00%	Testing other DL techniques to improve the classification accuracy and using other benchmark datasets with different skin disorders
[9]	2022	A DNN-based multi-modal classifier using wound images and their locations:- body map development, multi-modal network	Wound Diagnosis	developed (AZH DS), public DS (Medetec DS), and developed a mixed DS (AZHMT DS)	(AlexNet + MLP, AlexNet + LSTM, ResNet50 + MLP, VGG16 + LSTM)	Max. Acc. on mixed class: varies from 82.48 to 100% the max. acc. on wound-class varies from 72.95 to 97.12% in various experiments	Adding more modalities and more data, and using more specific wound image classifier and wound location classifier networks
[40]	2023	Mapping heterogeneous data features, fusion of clinical skin image & patient clinical data, feature extraction & attention mechanisms	Skin Cancer	PAD-UFES-20:- skin images and patient information	CNN: (VGGNet19, ResNet50, DenseNet121 & Inception-V3)	Achieved accuracy of 80.42% (an improvement of about 9% compared with the model accuracy using only medical images)	expanding scope of data collection: various types of imaging data (CT, MRI, US) and their corresponding clinical information

300 **5.1. Methods used for building diagnostic models for skin diseases**

301 In the final analysis of this systematic review, the nine studies identified proposed and
302 demonstrated the MMDF approach for the diagnosis of different skin diseases using their
303 corresponding datasets. The studies utilized different methods and algorithms that include CNN,
304 random forest, multilayer perceptron (MLP), long-short term memory (LSTM), the color
305 constancy algorithm, and hyperparameter optimization (HPO) algorithms. Accordingly, 88.9%
306 of the studies (8 articles) primarily utilized the CNN algorithm along with CNN architectures,
307 while 11.1% of the studies utilized MLP and LSTM along with CNN architectures including
308 ResNet50, VGG16, and AlexNet. In general, the studies employed different methods to
309 demonstrate the DL-based methods for combining different modalities of patient data using
310 different methods, such as the attention-based mechanism for combining images and metadata
311 features, a multimodal transformer using the Vision Transformer (ViT) model, and mapping
312 heterogeneous data features. In addition, DCNN architectures such as Densenet121, ILSVRC
313 2015, VGG16, VGGNet19, ResNet50, ResNet101, DenseNet121, Inception-V3, AlexNet with
314 MLP, AlexNet with LSTM, ResNet50 with MLP, and ViT models were utilized for feature
315 extraction and transfer learning purposes.

316 **5.2. Fusion strategies suggested for skin disease diagnosis**

317 Generally, data fusion techniques determine some issues, including the method of integrating
318 data, the data being fused or integrated, and the level at which data will be integrated. The
319 studies used for this review demonstrated various fusion approaches, mainly feature fusion,
320 model fusion, image fusion, and MMDF techniques. In this regard, 89% of the selected studies
321 analyzed in this review implemented MMDF approaches for integrating mainly clinical images
322 and textual medical data. Whereas only one study (11%) demonstrated the MMDF approach for
323 combining two imaging modalities (dermatoscopic and macroscopic images) with patient
324 metadata [35].

325 As reported by the studies used in this review, various fusion strategies have been experimented
326 with on a particular dataset while developing a diagnostic model for specific skin disease(s).
327 Accordingly, the fusion methods or strategies include integrating multiple imaging modalities (2

328 image modalities in this case) with textual patient data [35], using a multiplication-based fusion
329 approach (used to control data imbalance) [36], using the metadata processing block
330 (MetaBlock) for enhancing features extracted from the images throughout the classification [7],
331 other study used a naive combination of the patient data classifier module and a whole slide
332 image classifier module [8]. Furthermore, using a DNN that has two encoders for extracting
333 image features and textual features, a MMDF module with intra-modality self-attention and
334 inter-modality cross-attention capability was experimented with, and it was reported that the
335 model outperformed other fusion models [37]. On the other hand, a neural network with a
336 multimodal transformer consisting of two encoders for both images and metadata and one
337 decoder to fuse the multimodal information using the ViT model to extract image features, a soft
338 label encoder for the metadata, and a mutual attention block to fuse the different features [38]. In
339 another study, a fusion system was developed using four procedures consisting of preprocessing
340 the image and metadata, feature extraction using six pre-trained models, feature concatenation
341 (using CNN through convolutional, pooling, and auxiliary layers), and finally classification of
342 skin disease [39]. Similarly, the feature concatenation method was used to develop a wound
343 classifier multimodal network by concatenating the image classifier and location-based classifier
344 outputs [9]. Finally, a skin cancer diagnostic model was developed following three procedures,
345 including extracting features (skin images and patient clinical data using CNN architectures),
346 using the attention mechanism (for handling the multimodal features), and finally developing a
347 feature fusion model [40].

348 **5.3. Achievements of MMDF techniques in diagnosing skin diseases**

349 As stated by the studies reviewed, in developing diagnostic models using MMDF techniques for
350 skin diseases, various DL methods and algorithms were used, including CNN, Random Forest,
351 MLP, and LSTM. The algorithms achieved sufficiently higher performances in their respective
352 studies while being tested on a particular dataset. Consequently, it was confirmed that MMDF
353 techniques outperform traditional baseline diagnostic approaches [7][36]. Furthermore, the
354 majority of the studies reviewed reported that the disease classification models achieved
355 accuracy of more than 80% [8][9][35][38]. A study using a DNN with two encoders that
356 implement a multimodal fusion module with intra-modality self-attention and inter-modality
357 cross-attention reported an accuracy of 76.8% [37]. Similarly, another study used in this review

358 that used medical image analysis based on feature extraction, feature concatenation, and
359 classification or diagnosis methods reported 99.94% accuracy in the classification or diagnosis of
360 seven selected skin diseases. In general, as the analysis results show, MMDF techniques are
361 significantly improving classification accuracies. Therefore, the utilization of multimodal data
362 fusion techniques based on the deep learning methods, algorithms, and models in different
363 settings (such as an ensemble of two or more of those methods, algorithms, and models) is a
364 potential research area that needs further investigation, especially for the diagnosis of NTDs.

365 **6. Conclusion**

366 In this systematic review, articles were collected from seven major and reputed sources where
367 427 study papers were organized, classified, screened and selected to analyze the application of
368 DL-based diagnostic models using multimodal data fusion techniques for the diagnoses of skin
369 related NTDs. Although there are studies that demonstrate the utilization of DL methods for the
370 diagnoses of NTDs, no previous studies were found regarding the implementation of MMDF
371 methods for the diagnoses of NTDs. Similar studies using MMDF for the diagnoses of other skin
372 diseases, such as skin cancer, are reviewed to extract information about the implementation of
373 these methods. In doing so, the selected studies are analyzed using parameters such as research
374 approaches used, disease(s) selected for the study, the dataset used, algorithms used, the
375 performance achieved, and future directions suggested by the study. Accordingly, although all
376 the reviewed studies used diverse research methods and datasets based on their problem, DL-
377 based CNN algorithms were found to be by far the most frequently used algorithm by all studies
378 reviewed. In addition, DNN-based network architectures were widely utilized. In general, the
379 implementation of MMDF methods for the diagnosis of skin diseases significantly enhances the
380 diagnostic performances of models as per different studies reviewed, as confirmed in this review.
381 Hence, utilizing MMDF methods for the diagnoses of skin diseases, particularly for skin related
382 NTDs, would be paramount towards developing DL-based diagnostic models for NTDs.

383 **References**

384 [1] WHO, *Ending the neglect to attain the sustainable development goals: a road map for*
385 *neglected tropical diseases 2021–2030*. 2020.

- 386 [2] W.-Q. Page, “Neglected tropical diseases.” [https://www.who.int/news-room/questions-](https://www.who.int/news-room/questions-and-answers/item/neglected-tropical-diseases)
387 [and-answers/item/neglected-tropical-diseases](https://www.who.int/news-room/questions-and-answers/item/neglected-tropical-diseases) (accessed Jan. 15, 2023).
- 388 [3] World Health Organization, *Ending the neglect to attain the Sustainable Development*
389 *Goals: A rationale for continued investment in tackling neglected tropical diseases 2021–*
390 *2030*. 2022. [Online]. Available: <https://apps.who.int/iris/handle/10665/70809>
- 391 [4] A. A. Souza *et al.*, “Diagnostics and the neglected tropical diseases roadmap: Setting the
392 agenda for 2030,” *Trans. R. Soc. Trop. Med. Hyg.*, vol. 115, no. 2, pp. 129–135, 2021, doi:
393 10.1093/trstmh/traa118.
- 394 [5] S. G. Abdela *et al.*, “Looking for NTDs in the skin; an entry door for offering patient
395 centered holistic care,” *J. Infect. Dev. Ctries.*, vol. 14, no. 6.1, pp. 16S–21S, 2020, doi:
396 10.3855/jidc.11707.
- 397 [6] F. Afza, M. Sharif, M. A. Khan, U. Tariq, H. S. Yong, and J. Cha, “Multiclass Skin
398 Lesion Classification Using Hybrid Deep Features Selection and Extreme Learning
399 Machine,” *Sensors*, vol. 22, no. 3, 2022, doi: 10.3390/s22030799.
- 400 [7] A. G. C. Pacheco and R. A. Krohling, “An Attention-Based Mechanism to Combine
401 Images and Metadata in Deep Learning Models Applied to Skin Cancer Classification,”
402 *IEEE J. Biomed. Heal. Informatics*, vol. 25, no. 9, pp. 3554–3563, 2021, doi:
403 10.1109/JBHI.2021.3062002.
- 404 [8] J. Höhn *et al.*, “Combining CNN-based histologic whole slide image analysis and patient
405 data to improve skin cancer classification,” *Eur. J. Cancer*, vol. 149, pp. 94–101, 2021,
406 doi: 10.1016/j.ejca.2021.02.032.
- 407 [9] D. M. Anisuzzaman, Y. Patel, B. Rostami, J. Niezgoda, S. Gopalakrishnan, and Z. Yu,
408 “Multi-modal wound classification using wound image and location by deep neural
409 network,” *Sci. Rep.*, vol. 12, no. 1, pp. 1–20, 2022, doi: 10.1038/s41598-022-21813-0.
- 410 [10] W. feng Shen, H. wei Tang, J. bo Li, X. Li, and S. Chen, “Multimodal data fusion for
411 supervised learning-based identification of USP7 inhibitors: a systematic comparison,” *J.*
412 *Cheminform.*, vol. 15, no. 1, pp. 1–16, 2023, doi: 10.1186/s13321-022-00675-8.
- 413 [11] J. Gao, P. Li, Z. Chen, and J. Zhang, “A survey on deep learning for multimodal data

- 414 fusion,” *Neural Comput.*, vol. 32, no. 5, pp. 829–864, 2020, doi: 10.1162/neco_a_01273.
- 415 [12] J. Lipkova *et al.*, “Artificial intelligence for multimodal data integration in oncology,”
416 *Cancer Cell*, vol. 40, no. 10, pp. 1095–1110, 2022, doi: 10.1016/j.ccell.2022.09.012.
- 417 [13] N. Luo, X. Zhong, L. Su, Z. Cheng, W. Ma, and P. Hao, “Artificial intelligence-assisted
418 dermatology diagnosis: From unimodal to multimodal,” *Comput. Biol. Med.*, vol. 165, no.
419 July, p. 107413, 2023, doi: 10.1016/j.compbimed.2023.107413.
- 420 [14] R. Beesetty *et al.*, “Leprosy Skin Lesion Detection: An AI Approach Using Few Shot
421 Learning in a Small Clinical Dataset,” *Indian J Lepr*, vol. 2023, pp. 89–102, 2023,
422 [Online]. Available: <http://www.ijl.org.in>
- 423 [15] Z. Ali *et al.*, “A Proposed Framework for Early Prediction of Schistosomiasis,”
424 *Diagnostics*, vol. 12, no. 12, pp. 1–25, 2022, doi: 10.3390/diagnostics12123138.
- 425 [16] N. Steyve, P. Steve, M. Ghislain, S. Ndjakomo, and E. pierre, “Optimized real-time
426 diagnosis of neglected tropical diseases by automatic recognition of skin lesions,”
427 *Informatics Med. Unlocked*, vol. 33, no. September, 2022, doi:
428 10.1016/j.imu.2022.101078.
- 429 [17] R. R. Yotsu, Z. Ding, J. Hamm, and R. E. Blanton, “Deep learning for AI-based diagnosis
430 of skin-related neglected tropical diseases: A pilot study,” *PLoS Negl. Trop. Dis.*, vol. 17,
431 no. 8, pp. 1–12, 2023, doi: 10.1371/journal.pntd.0011230.
- 432 [18] F. Castanedo, “A review of data fusion techniques,” *Sci. World J.*, vol. 2013, 2013, doi:
433 10.1155/2013/704504.
- 434 [19] N. J. Shoumy, L. M. Ang, K. P. Seng, D. M. M. Rahaman, and T. Zia, “Multimodal big
435 data affective analytics: A comprehensive survey using text, audio, visual and
436 physiological signals,” *J. Netw. Comput. Appl.*, vol. 149, p. 102447, 2020, doi:
437 10.1016/j.jnca.2019.102447.
- 438 [20] T. Zhang *et al.*, “A Feature Fusion Method with Guided Training for Classification
439 Tasks,” *Comput. Intell. Neurosci.*, vol. 2021, no. c, 2021, doi: 10.1155/2021/6647220.
- 440 [21] X. Lu, X. Duan, X. Mao, Y. Li, and X. Zhang, “Feature Extraction and Fusion Using

- 441 Deep Convolutional Neural Networks for Face Detection,” *Math. Probl. Eng.*, vol. 2017,
442 2017, doi: 10.1155/2017/1376726.
- 443 [22] N. AlDahoul, H. Abdul Karim, and A. S. Ba Wazir, “Model fusion of deep neural
444 networks for anomaly detection,” *J. Big Data*, vol. 8, no. 1, 2021, doi: 10.1186/s40537-
445 021-00496-w.
- 446 [23] H. Kaur, D. Koundal, and V. Kadyan, “Image Fusion Techniques: A Survey,” *Arch.*
447 *Comput. Methods Eng.*, vol. 28, no. 7, pp. 4425–4447, 2021, doi: 10.1007/s11831-021-
448 09540-7.
- 449 [24] Y. Wang, J. Cai, D. C. Louie, Z. J. Wang, and T. K. Lee, “Incorporating clinical
450 knowledge with constrained classifier chain into a multimodal deep network for
451 melanoma detection,” *Comput. Biol. Med.*, vol. 137, no. June, p. 104812, 2021, doi:
452 10.1016/j.combiomed.2021.104812.
- 453 [25] I. An, M. Harman, and I. Ibiloglu, “Topical Ciclopirox Olamine 1%: Revisiting a Unique
454 Antifungal,” *Indian Dermatol. Online J.*, vol. 10, no. 4, pp. 481–485, 2017, doi:
455 10.4103/idoj.IDOJ.
- 456 [26] O. Yélamos, L. Mary Diem, R. P. Braun, K. K. French, and A. A. Marghoob, *Dermoscopy*
457 *for Dermatopathologists*. Elsevier Inc., 2019. doi: 10.1016/B978-0-323-37457-6.00028-6.
- 458 [27] M. Pawłowski, A. Wróblewska, and S. Sysko-Romańczuk, “Effective Techniques for
459 Multimodal Data Fusion: A Comparative Analysis,” *Sensors*, vol. 23, no. 5, pp. 1–16,
460 2023, doi: 10.3390/s23052381.
- 461 [28] M. J. Page *et al.*, “The prisma 2020 statement: An updated guideline for reporting
462 systematic reviews,” *Med. Flum.*, vol. 57, no. 4, pp. 444–465, 2021, doi:
463 10.21860/medflum2021_264903.
- 464 [29] PRISMA, “PRISMA: TRANSPARENT REPORTING of SYSTEMATIC REVIEWS and
465 META-ANALYSES.” <http://www.prisma-statement.org/> (accessed Nov. 15, 2023).
- 466 [30] R. AI, “RAYYAN.” <https://www.rayyan.ai/> (accessed Oct. 20, 2023).
- 467 [31] I. Bakkouri and K. Afdel, “Computer-aided diagnosis (CAD) system based on multi-layer

- 468 feature fusion network for skin lesion recognition in dermoscopy images,” *Multimed.*
469 *Tools Appl.*, vol. 79, no. 29–30, pp. 20483–20518, 2020, doi: 10.1007/s11042-019-07988-
470 1.
- 471 [32] R. Mahum and S. Aladhadh, “Skin Lesion Detection Using Hand-Crafted and DL-Based
472 Features Fusion and LSTM,” *Diagnostics*, vol. 12, no. 12, 2022, doi:
473 10.3390/diagnostics12122974.
- 474 [33] D. Zhang, A. Li, W. Wu, L. Yu, X. Kang, and X. Huo, “CR-Conformer: a fusion network
475 for clinical skin lesion classification,” *Med. Biol. Eng. Comput.*, no. 0123456789, 2023,
476 doi: 10.1007/s11517-023-02904-0.
- 477 [34] S. Afra and R. Alhajj, “al Pr p ro of,” *Physica A*, p. 123151, 2019, doi:
478 10.1016/j.jpi.2023.100341.
- 479 [35] J. Yap, W. Yolland, and P. Tschandl, “Multimodal skin lesion classification using deep
480 learning,” *Exp. Dermatol.*, vol. 27, no. 11, pp. 1261–1267, 2018, doi: 10.1111/exd.13777.
- 481 [36] W. Li, J. Zhuang, R. Wang, and J. Zhang, “FUSING METADATA AND
482 DERMOSCOPY IMAGES FOR SKIN DISEASE DIAGNOSIS School of Data and
483 Computer Science , Sun Yat-sen University , China Key Laboratory of Machine
484 Intelligence and Advanced Computing , MOE , Guangzhou , China Department of
485 Computer Science an,” pp. 1996–2000, 2020.
- 486 [37] C. Ou *et al.*, “A deep learning based multimodal fusion model for skin lesion diagnosis
487 using smartphone collected clinical images and metadata,” *Front. Surg.*, vol. 9, no.
488 October, pp. 1–9, 2022, doi: 10.3389/fsurg.2022.1029991.
- 489 [38] G. Cai, Y. Zhu, Y. Wu, X. Jiang, J. Ye, and D. Yang, “A multimodal transformer to fuse
490 images and metadata for skin disease classification,” *Vis. Comput.*, vol. 39, no. 7, pp.
491 2781–2793, 2023, doi: 10.1007/s00371-022-02492-4.
- 492 [39] S. N. Almuayqil, S. Abd El-Ghany, and M. Elmogy, “Computer-Aided Diagnosis for
493 Early Signs of Skin Diseases Using Multi Types Feature Fusion Based on a Hybrid Deep
494 Learning Model,” *Electron.*, vol. 11, no. 23, 2022, doi: 10.3390/electronics11234009.
- 495 [40] Q. Chen, M. Li, C. Chen, P. Zhou, X. Lv, and C. Chen, “MDFNet: application of

496 multimodal fusion method based on skin image and clinical data to skin cancer
497 classification,” *J. Cancer Res. Clin. Oncol.*, vol. 149, no. 7, pp. 3287–3299, 2023, doi:
498 10.1007/s00432-022-04180-1.