User Awareness System to Diagnose Dermatological Diseases

Article *in* International Journal of Computer Applications · December 2020

6 authors, including:

Vithushiyan Pathivarathan
Sri Lanka Institute of Information Technology
2 PUBLICATIONS **0** CITATIONS

Philipreman Kasvithan
Sri Lanka Institute of Information Technology
2 PUBLICATIONS **0** CITATIONS

Sinmayan Gunasekaran
Sri Lanka Institute of Information Technology
2 PUBLICATIONS **0** CITATIONS

Naveenan Thavabalasingham
Sri Lanka Institute of Information Technology
2 PUBLICATIONS **0** CITATIONS

User Awareness System to Diagnose Dermatological Diseases

Vithushiyan Pathivarathan
Department of Information Technology Sri Lanka Institute of Information Technology
Malabe, Sri Lanka

Naveenan Thavabalasingham
Department of Information Technology Sri Lanka Institute of Information Technology
Malabe, Sri Lanka

Kasvithan Philipreman
Department of Information Technology Sri Lanka Institute of Information Technology
Malabe, Sri Lanka

Sinmayan Gunasekaran
Department of Information Technology Sri Lanka Institute of Information Technology
Malabe, Sri Lanka

Sanjeevi Chandrasiri
Department of Information Technology Sri Lanka Institute of Information Technology
Malabe, Sri Lanka

Thilini Weerasooriya
Department of Information Technology Sri Lanka Institute of Information Technology
Malabe, Sri Lanka

ABSTRACT
Nowadays, humans' health is deteriorating by dermatological diseases, and the spreading rate is high. Most people are not aware of skin diseases. As they do not realize these diseases' seriousness, they try to treat with some remedies by themselves, even without knowing what the actual disease is. Nevertheless, it is not a suitable way to cure the disease, leading to future complications. So still the dermatological diseases remain as one of the main categories of common health issues. A few people prefer to use computerized systems to evaluate the disease conditions these days. Moreover, it is essential to know about the diseases to manage that condition and prevent escalation. Therefore, the proposed system is implemented to give users some knowledge about dermatological diseases as much as possible. The users can get awareness and predict skin diseases and complications from the data mining technique. The user can identify the stage of the dermatological disease by applying the classification algorithm. Furthermore, this system will also scrap web pages related to that disease from known or system verified websites. The content analysis is based on the machine learning process, especially using Neural Language Processing. Hence, the system will undeniably be useful to the users to summarize skin diseases and get concerns from a dermatologist.

General Terms
Convolutional Neural Networks, TensorFlow, Application Programming Interface,

Keywords
Dermatological diseases, Image processing, Data mining, Web scraping, Natural Language Processing

1. INTRODUCTION
The infection rate of dermatological disease is higher than the other diseases because the skin is the largest organ in the human body. It is assumed that 20-25% of the global human population is affected by skin infection with a constant increase [1]. Skin diseases are transmitted through direct person-to-person contact or any other pest. Many counties focus on making dermatological disease awareness programs to prevent humans from getting affected by skin diseases [2]. Knowing the condition of the dermatological disease will help to cure it or prevent it from escalating.

Especially in developing countries, automated system detection of dermatology impacts both patients and dermatologists is essential. People can quickly get rid of being affected, or even if they get affected, they can quickly cure most of the dermatological diseases if the diseases are detected at the earlier Stage.

A facility in this application, to create awareness about the dermatological diseases, the user has to input some attributes that the system will use to diagnose the diseases' category. This application will notify the user about the disease identified by analyzing the websites through web scraping globally. Therefore, the user can be aware of that particular disease that is being identified recently.

There are some mobile applications to identify the diseases type by capturing the affected skin's image. The level of accuracy of the detection of dermatological diseases is deficient. Therefore, a classification model with high accuracy is developed in this application. Most peoples do not become aware of these kinds of diseases. For that purpose, functionality is added to identify the Stage of dermatological diseases. The users can get a clear view of the diseases and the Stages of disease they are by using this application. This will help the patients to cure immediately.

2. LITERATURE REVIEW
Various parties have done several types of research regarding dermatological disease analysis. Most of the researches is based on image processing. These researches are focused on the accuracy of image processing. In Bangladesh, one research was found based on two phases- first, pre-process the color skin images to extract significant features and later identify the diseases [3].

This system has detected nine types of skin diseases, including Acne, Leprosy, and Vitiligo. The accuracy rate for disease identification is 90%. Another research was found in the United Kingdom is also focusing on image processing. The skin images have been analyzed simple graphic user interface and mobile neural network. Easy access control of the integrated camera was allowed by Application Programming Interface (API). The image which is input is presented in a dimensional vector contains Red, Green, and Blue (RGB) values.

Some other researches are based on low-cost smartphone systems. The researchers have proposed a low-cost smartphone-based intelligent scheme in the United Kingdom research that allows people for regular skin examinations [4-6]. Several technologies have been used in that system, for instance, using an inexpensive mobile device to capture high-resolution skin images using artificial Neural Networks for local abnormal/standard skin image classification, i.e., distinguishing between typical and disease-related skin images [7-9].

In a research found in Bangladesh, the system they proposed works on two dependent steps - the first detects skin anomalies and identifies the diseases. The system operates on visual input, i.e., high-resolution color images and patient history. The system uses color image processing methods, k-means clustering, and color gradient methods to identify the diseased skin in machine intervention. For disease classification, the system possibilities to feedforward backpropagation artificial neural networks. The system displays a diseased skin detection accuracy of 95.99% and disease identification accuracy of 94.016% while tested for a total of 2055 diseased ranges in 704 skin images for six diseases.

In some models, skin disease is identified by evaluating skin disease images by using the grey normalized symmetrical simultaneous occurrence stencils (GLCM) method. These systems work with relational databases to the storage of implying the need for textual skin images. These systems can also work for some images directly over feature vectors. However, in this model, only the skin texture will be analyzed [10-12].

A survey-based on web news retrieval and mining clearly explained the identification of dermatological diseases using web mining. The system they proposed is mainly using the web mining technique. Web mining is mining the news contents by data mining and text mining techniques. The news retrieval process starts with user input. These news retrieval systems try to reduce the retrieval domain and filtering result. News or articles are divided into structured data, and those data compare news similarities for ranking the news. News filtering checks noise and spam news and filters the redundant low informative news based on word similarities and aligns news extraction performed by automatically extracting news parts in a structured manner.

These extractions are divided into layout-dependent extraction and independent extraction, and visual-featured-based methods (for differentiated noise content). The news content analysis part consists of three main fields of research. Pattern mining is performed on a corpus of news articles to find meaningful patterns: opinion Mining and Sentiment Analysis in the news, performed on the document level—furthermore, Topic Detection and Tracking, which is performed on the whole web [13].

3. METHODOLOGY

The Proposed system will be carried out under four components.

A. Diagnosing Dermatological Disease

B. Classify the type of Dermatological Disease

C. Identify the Dermatological Diseases Stages

D. Notify the user about Dermatological Diseases

3.1 Diagnosing Dermatological Disease

Dermatological diseases are widespread nowadays. The image processing technique is used to identify skin diseases. The user has to insert the image of the affected area. Then the system will let the user know which kind of dermatological disease is that. In that respect, the following techniques are used to do image processing techniques.

3.1.1 Image Acquisition

Image acquisition is the initial step of the image processing technique. In this step, the images of the dermatological diseases are captured through the camera of the smartphone. Disease images are shown in Figure 1

3.1.2 Image Pre-processing

Image pre-processing is the foremost step in the image processing technique. In this method, the images' noises, such as hair, clothing, and, other artifacts, will be removed. The image processing technique's primary purpose is to improve skin diseases' quality by removing the related parts.

Gaussian filter is highly suggested as it removes the speckle noise [14]. Gaussian kernel coefficients are sampled from the 2D Gaussian function as shown in equation (1)

$$(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \qquad (1)$$

3.1.3 Image Segmentation

Image segmentation is an important technique used in image processing techniques. This technique is used to determine the shape and size of the border of the image. It separates the objects from the background based on different features. The segmented image is shown in Figure 2.

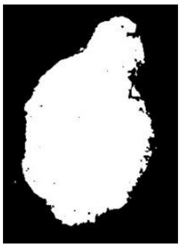

Figure 2: Segmented image

3.1.4 Feature Extraction

The system users can quickly get information about the dermatological disease they have been affected by extracting information from the diseases' images. The targeted areas' features are extracted using techniques feature extraction methods are Skewness Asymmetric degree of pixels distribution in the specified window around its mean, as in equation (2), (3).

$$\boldsymbol{m = \frac{1}{mn} \sum_{i=1}^{m} \sum_{j=1}^{n} p(i, j)} \qquad (2)$$

$$\boldsymbol{m = \frac{1}{mn} \sum_{i=1}^{m} \sum_{j=1}^{n} \left(\frac{p(i,j) - m}{\sigma}\right)^3} \qquad (3)$$

Different intensity values between the region pixels are shown in equation (4).

$$\boldsymbol{c = \sum_{i,j} (i - j)^2 p(i, j)} \qquad (4)$$

From a small region which is centered on the pixel, the features can be extracted. Due to the pixel's potent discrimination ability, the image texture, an essential factor of the product surface for recognition, can be extracted as a pixel feature

3.1.5 Image Classification

Image classification is the final process used in the image processing technique. It is the process of selected features, which are used to identify and classify dermatological diseases. Here the CNN (Convolutional Neural Network), a supervised machine learning algorithm, will be used to identify the affected area. For handling the CNN, TensorFlow and Keras were used. For diagnosing the diseases, the trained model was used in the mobile application. Because of the TensorFlow model's heaviness model, the applications' functions and the whole application will be slow. To overcome the issue TF model was converted into the TensorFlow Lite model and used the TensorFlow Lite model in the mobile application.

3.2 Classify the type of Dermatological Disease

They are classifying the type of erythemato-squamous skin disease. In which there are six different categories.

C1: psoriasis

C2: seboreic dermatitis

C3: lichen planus

C4: pityriasis rosea

C5: cronic dermatitis

C6: pityriasis rubra pilaris

Almost every category contains standard features. It is challenging to classify the category of disease. Here system uses a machine learning technique to identify the category of diseases using its characteristics.

3.2.1 Data collection

The data has been taken from the database of the UCI machine repository.

3.2.2 Data pre-processing

The database's collected data is not clean as it consists of noisy, incorrect, missing values. The system removed the rows with missing values.

3.2.3 Model building for classification

There are several algorithms to perform classification [15]. Here use two main algorithms to classify the erthemato-squamous [16] disease. Algorithms used are gradient boosting and naïve Bayes to classify the category of disease. Naïve Bayes classifier works [17] by merely assuming that all features are independent. as in equation (5), (6)

$$P(c \mid x) = \frac{P(x \mid c)P(c)}{P(x)} \qquad (5)$$

$$P(c \mid X) = P(x_1 \mid c) \times P(x_2 \mid c) \times \cdots \times P(x_n \mid c) \times P(c) \qquad (6)$$

3.2.4 Evaluation

The dermatology dataset is partitioned into two parts: training the model (80%) and testing the model (20%).

3.3 Identify the Dermatological Diseases Stages

This part represents the implementation of identifying the level of dermatological disease by applying the classification algorithm. By considering the features extracted from the image, are used to classify the image into various classes based on different characteristics and predict the time taking to reach the critical Stages. The system typically following a procedure that involves four main steps.

3.3.1 Image Acquisition

The first Stage of this system is the image acquisition stage. After the image has been obtained, several processing methods can be applied to the image to perform the various vision tasks required today. However, if the image has not been acquired acceptably, the intended tasks may not be achievable, even with some form of image enhancement. The non-visual data of patient disease Images had to be collected.

3.3.2 Feature Extraction

Feature extraction is a technique used for collecting several features from images and dimensionality reduction. Feature extraction methods are applied to get features that will be useful in classifying and recognition images.

3.3.3 Image Classification

The unique features of the enhanced images were extracted using Different methods in the feature extraction step. Based on the features, the images were classified as diseased skin and normal skin. Image classification analyses the numerical properties of several image features and organizes data into categories. Classification algorithms typically employ two phases of processing, like training and testing. In the initial training phase, characteristic properties of typical image features are isolated, and, based on these, a unique description of each classification category, a training class, is created. In the subsequent testing phase, these feature-space partitions are used to classify image features [18].

Convolutional neural network (CNN) in deep learning [19] is the technique used to Classifying the affected level of dermatological disease. The pre-trained model is available in VGG19 models used to do the Classifying the affected level of dermatological disease. Keras provides a simple and modular API to create a trained neural network and run on top of the Tensor Flow library, the popular deep learning models [20]. These pre-trained models are trained on a large dataset in the Image Net consist of 3000 training images and 200 testing images.

The image dimension is 32 x 32 x 3 with six classes. Deep learning models require a large amount of data to make an accurate prediction.

The image data augmentation technique is used to expand the size of a training dataset artificially. Data augmentation feature supported in Keras deep learning library via the Image Data Generator [21] class generates the input image variations changing the size, orientation, and shifting the image (left, right, top, bottom). The system trains the dataset using pre-trained models and get the accuracy and the model and then get the highest accuracy to do the Classifying the affected level recognition process.

VGG-19 is a CNN that is 19 layers deep. The system can load a pre-trained version of the network trained on more than a million images. As a result, the network has learned rich feature representations for an extensive range of images. VGG19 based model was performed with a test accuracy of 78.12%.

3.4 Notify the user about Dermatological Diseases

This part is aware of dermatological disease users using news, publication, and more other reliable source. Using a search engine and web scrape method gathers web site related to the dermatological disease. The central part of it is to get more news from different websites. Using Natural Language Processing (NLP) technique, compare and analyze, and remove ambiguous content.

3.4.1 Web scraping and Web content extraction

The aim of the project not to scrape all news websites. Before extract content from a webpage, the user wants to know where the user wants to get it. Related News can be getting by the google hacking technique. The name is a little confusing, but this is not non-ethical. Using this technique, the user can get a website with a condition. Users can input crafted text into search pages; it will return the expected news.

- Intitle = search result by page title
- Intext = search result by the text of page only
- Date range = search result by the search range

Furthermore, there are many options like result by author name, site, filetype, and more [22]. Web news extraction is not a difficult part for a well-structured website. Most of the websites will struct their news into paragraph tags. Nevertheless, paragraph tags can be with more unwanted or other words.

3.4.2 Content Analysis

Unwanted and content that is not related to the news want to be removed. Using comparison, the user can identify critical data and remove unwanted. However, before analysis, the content wants to be separated as sentences, then it wants to be changed into another form because machines cannot compare words by any language.

3.4.3 Word Embedding

Word embedding is the process that changes words or phrases into a vector of numbers that model with language and feature learning. It is not like a word to ASCII or UNICODE, which will only present a word with just letters, but word embedding can present the meaning or emotion of words.

```
['There is a person behind tree', 'The person is in behind the tree']
tf.Tensor(
[[-0.0393908   0.01541872  0.00575714 ...  0.01080611 -0.02330867
  -0.03412762]
 [-0.01387427  0.08217918 -0.02722782 ... -0.01692257  0.00868729
  -0.00553428]], shape=(2, 512), dtype=float32)
```

Figure 3: Sentence to vector using word embedded method

high-quality embedding vectors, a large amount of training data is necessary [23].

Word2vec is a model that uses word embedding. The basic model of word2vec is accurate enough to be applied to similarity analysis [24]. So, using the word2vec system will analyze using two types of analysis.

- Sentiment analysis
- Semantic analysis

3.4.4 Sentiment Analysis

Sentiment analysis uses vectors; it will identify a sentence's emotion using it indicates either positive or negative. Naïve Bayes algorithm is a better algorithm for accuracy than another algorithm [25].

3.4.5 Semantic Analysis

Semantic analysis is modern and complex than the sentiment analysis algorithm—a much more statically reliable approach. Analyze the vector, and it will compare sentence by meaning.

Another direction of work presumes automatic extraction of results in publication, describing these results efficiency and areas of practical applications [26].

The similarity comparison with some words can show in numbers that maximum as one and minimum as zero. The collective result will be shown as a heatmap in Figure 4

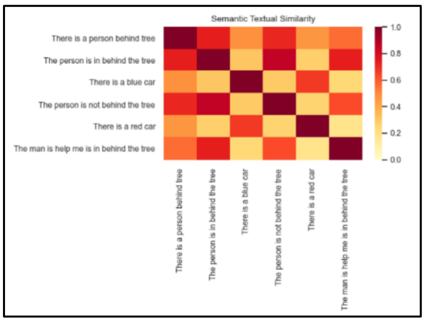

Figure 4: Sentence Semantic comparison in heatmap

4. RESULTS & DISCUSSIONS

The proposed system can successfully diagnose Melanoma & Psoriasis with 90% accuracy. After the user uploaded the affected skin image, the mobile application will process the image and display the name, symptoms, and treatment shown in Figure 5 as the output.

Figure 5: Output of Disease Diagnosing

The disease's diagnosis was tested against 40 sample inputs, and out of the samples, 37 inputs were accurate. Thus, the expected accuracy rate is more than 90%. The following Table 1 illustrates the accuracy rate of the diseases.

Table 1. Accuracy of the Diseases

Disease	*Total Images*	*Disease Diagnose*	*Accuracy (%)*
Melanoma	20	18	90%
Psoriasis	20	19	95%

When a user inputs the attributes, the system will display the category of the disease belongs to. Here two algorithms were used to classify the disease. The following Table 2 shows the accuracy rates of the algorithms.

Table 2. Accuracy of the Algorithm

Algorithm	*Accuracy (%)*
gradient boosting	95.8
naïve bayes	88.8

The system has successfully diagnosed the stages of Nevus & Melanoma as Begin or Malignant. The system will help the user identify the disease, and its stage user will decide whether it can be harmful or not or any other way to cure it, and it helps the user identify skin diseases in the early stages. Identified Stage is shown in Figure 6

Figure 6: Identified Stage

The system was tested against 80 sample inputs, and out of the samples, 68 inputs were accurate. Thus, the expected accuracy rate is moreover than 85%. The following Table 3 shows the accuracy rate of the stages.

Table 3. Accuracy of the Diseases Stage

Disease	*Disease Image*	*Diagnose Stage*	*Stage*	*Accuracy (%)*
Melanoma	20	17	Begin	85
	20	16	Malignant	80
Nevus	20	18	Begin	90
	20	17	Malignant	85

There is also a facility to notify the users about the details of the widespread dermatological diseases. Use can get the latest news update about the dermatological as soon in the mobile application.

5. CONCLUSION

This research project aims to provide an efficient and effective mobile application and awareness to the people. All the components are developed with user-friendliness. The collection of data is very accurate and tested many times. This proposed mobile application will provide the most reliable and accurate details of the diseases to the users.

6. FUTURE WORK

In the future, the application will be modified to diagnose many dermatological diseases with high efficiency. The application will be developed on many platforms, including iOS, with the same features. Moreover, many languages, including Tamil and Sinhala, will be used in upcoming versions.

7. ACKNOWLEDGEMENT

This research project was supported by the supervisor and the co-supervisor of this project. We want to express our gratitude to them. Furthermore, we thank all the people who helped and provided information to finish the project with satisfaction.

8. REFERENCES

[1] Ulf Mader, Niko Quiskamp, Soren Wildenhain, Thomas Schmidts, Peter Mayser, Frank Runkel and Martin Fiebich, , Image-Processing Scheme to Detect Superficial Fungal Infections of the Skin, Germany: Hindawi, 2015

[2] Beth Snyder Bulik, 'Almirall 'Shared Skin' dermatology

disease awareness campaign gets employees involved', Jun 13, 2016 [online] Available: https://www.fiercepharma.com/marketing/almirall-shared-skin-dermatology-disease-awareness-campaign-gets-employees-involved

[3] I. Rahat Yasir,,* Md. Ashiqur Rahman, and Nova Ahmed, Electronic and Computer Engineering Department, North South University, Dhaka, Bangladesh." Dermatological Disease Detection using Image Processing and Artificial Neural Network"- 8th International Conference on Electrical and Computer Engineering 20-22 December, 2014, Dhaka, Bangladesh

[4] Abderrahim Bourouisa ,Ali Zerdazia , Mohammed Fehamb , Abdelhamid Bouchachiac* "M-Health: Skin Disease Analysis System using Smart phone's camera"- The 8th International Symposium on Intelligent Systems Techniques for Ad hoc and Wireless Sensor Networks (IST-AWSN).

[5] Abderrahim Bourouis, Mohamed Feham,Abdelhamid Bouchachia. Ubiquitous Mobile Health Monitoring System for Elderly (UMHMSE). International Journal of Computer Science & Information Technology ; June 2011 ,Vol 3, No 3.

[6] Al-Taee M.A , Jaradat N.A , Ali, D.M.A. Mobile phone-based health data acquisition system using Bluetooth technology .Applied Electrical Engineering and Computing Technologies (AEECT); 2011 ,p.1-6.

[7] Qing Pan, Pan Yang, Rui Zhang, Chengyu Lin, Shijin Gong, Li Li, Jing Yan, Gangmin Ning . A mobile health system design for home and community use, Biomedical and Health Informatics (BHI), 2012 IEEE-EMBS International Conference ;2012 , p. 116 - 119 .

[8] Asada H, Shaltis P, Reisner A, Rhee S, Hutchinson R. Mobile monitoring with wearable photoplethysmograhic biosensors. IEEE Eng Med Biology Mag ;May-June 2003,p. 28 - 40 .

[9] Ross P. Managing care through air. IEEE Spectrum; Dec 2004, p.26- 31.

[10] Anal Kumar MittraandDr. Ranjan Parekh - Automated Detection of Skin Diseases Using Texture Features.

[11] Fitzpatrick, T.B., Rhodes, A.J., Sober, A., Mihm, M. (1988), Primary malignant melanoma of the skin: The call for action to identify persons at risk, to discover precursor lesion, to detect early melanoma. Pigm Cell Res 9(1): 110-7

[12] Granot, Y., Ivorra, A., Rubinsky, B. (2008), A New Concept for Medical Imaging Centered on Cellular Phone Technology PLoS ONE 3(4): e2075. doi:10.1371/journal.pone.0002075

[13] Roya Hassanian-esfahani, Mohammad-javad Kargar, " A Survey on Web News Retrieval and Mining", 2008 9th International Conference on Computer-Aided Industrial Design and Conceptual Design.

[14] Muhammad Qasim Khan,Ayyaz Hussain,Saeed Ur Rehman,Umair Khan,"Classification of Melanoma and Nevus in Digital Images for Diagnosis of Skin Cancer",Department of Computer Science and Software Engineering, International Islamic University, Islamabad 44000, Pakistan

[15] Han J, Kamber M. Data mining : concepts and techniques. 3rd ed. Burlington, MA: Elsevier; 2011.

[16] Güvenir HA, Demiröz G, Ilter N. Learning differential diagnosis of erythemato-squamous diseases using voting feature intervals. Artificial Intelligence in Medicine. 1998;13(3):147–65.

[17] Danjuma K, Osofisan AO. Evaluation of Predictive Data Mining Algorithms in Erythemato-Squamous Disease Diagnosis. International Journal of Computer Science Issues. 2014;11(6):85–94.

[18] Damilola A. Okuboyejo, Oludayo O. Olugbara, and Solomon A. Odunaike (2013) - Automating Skin Disease Diagnosis Using Image Classification-Proceedings of the World Congress on Engineering and Computer Science 2013, Volume II, San Francisco, USA.

[19] S.Saha,"towardsdatascience,"[Online].Available:https://towardsdatascience.com/a-comprehensive-guide-toconvolutional-neural-networks-the-eli5-way-3bd2b1164a53. [Accessed 15 December 2018].

[20] F. Chollet, "Building powerful image classification models using very littledata,"2016.[Online].Available: https://blog.keras.io/buildingpowerful-image-classificationmodels-using-very-little-data.html. [Accessed: 05-Mar-2019]

[21] S. Paranavitana, Inscriptions of Ceylon, Department of Archaeology, Colombo, Sri lanka, 1970.

[22] Esteban Borges, "Exploring Google Hacking Techniques", [online], Available: https://securitytrails.com/blog/google-hacking-techniques

[23] Young-Jun Lee, Chan-Yong Park, Ho-Jin Choi, "Word-level Emotion embedding based on Semi-Supervised Learning for Emotional Classification in Dialogue", Korea Electric Power Corporation, Grant number: R18XA05

[24] Chunyu Xia, Tieke He, Wenlong Li, Zemin Qin, Zhipeng Zou, "Similarity Analysis of Law Documents Based on Word2vec", IEEE 19th International Conference on Software Quality, Reliability and Security Companion (QRS-C), 2019

[25] Satuluri Vanaja, Meena Belwal, "Aspect-Level Sentiment Analysis on E-Commerce Data", Proceedings of the International Conference on Inventive Research in Computing Applications (ICIRCA), 2018.

[26] Gennady Osipov, Ivan Smirnov, Ilya Tikhomirov, Olga Vybornova, "Technologies for Semantic Analysis of Scientific Publications", Institute for Systems Analysis Russian Academy of Sciences, 2012