Comet Assay Classification for Buccal Mucosa's DNA Damage Measurement with Super Tiny Dataset Using Transfer Learning

Chapter · January 2020

DOI: 10.1007/978-3-030-14132-5_22

Studies in Computational Intelligence 830

Maciej Huk
Marcin Maleszka
Edward Szczerbicki *Editors*

Intelligent Information and Database Systems: Recent Developments

Springer

Studies in Computational Intelligence

Volume 830

Series editor

Janusz Kacprzyk, Polish Academy of Sciences, Warsaw, Poland
e-mail: kacprzyk@ibspan.waw.pl

The series "Studies in Computational Intelligence" (SCI) publishes new developments and advances in the various areas of computational intelligence—quickly and with a high quality. The intent is to cover the theory, applications, and design methods of computational intelligence, as embedded in the fields of engineering, computer science, physics and life sciences, as well as the methodologies behind them. The series contains monographs, lecture notes and edited volumes in computational intelligence spanning the areas of neural networks, connectionist systems, genetic algorithms, evolutionary computation, artificial intelligence, cellular automata, self-organizing systems, soft computing, fuzzy systems, and hybrid intelligent systems. Of particular value to both the contributors and the readership are the short publication timeframe and the world-wide distribution, which enable both wide and rapid dissemination of research output.

The books of this series are submitted to indexing to Web of Science, EI-Compendex, DBLP, SCOPUS, Google Scholar and Springerlink.

More information about this series at http://www.springer.com/series/7092

Maciej Huk · Marcin Maleszka ·
Edward Szczerbicki
Editors

Intelligent Information and Database Systems: Recent Developments

Springer

Editors
Maciej Huk
Department of Information Systems
Wrocław University of Science and Technology
Wrocław, Poland

Marcin Maleszka
Department of Information Systems
Wrocław University of Science and Technology
Wrocław, Poland

Edward Szczerbicki
Department of Management
Gdańsk University of Technology
Gdańsk, Poland

ISSN 1860-949X ISSN 1860-9503 (electronic)
Studies in Computational Intelligence
ISBN 978-3-030-14131-8 ISBN 978-3-030-14132-5 (eBook)
https://doi.org/10.1007/978-3-030-14132-5

Library of Congress Control Number: 2019932697

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Program Committee

Ajith Abraham, Machine Intelligence Research Labs, USA
Muhammad Abulaish, South Asian University, India
Andy Adamatzky, University of the West of England, UK
Waseem Ahmad, Waiariki Institute of Technology, New Zealand
Bashar Al-Shboul, University of Jordan, Jordan
Lionel Amodeo, University of Technology of Troyes, France
Toni Anwar, Universiti Teknologi Malaysia, Malaysia
Tsuyoshi Arai, Okayama Prefectural University, Japan
Taha Arbaoui, University of Technology of Troyes, France
Ahmad Taher Azar, Benha University, Egypt
Thomas Bäck, Leiden University, The Netherlands
Amelia Badica, University of Craiova, Romania
Costin Badica, University of Craiova, Romania
Kambiz Badie, ICT Research Institute, Iran
Zbigniew Banaszak, Warsaw University of Technology, Poland
Dariusz Barbucha, Gdynia Maritime University, Poland
Ramazan Bayindir, Gazi University, Turkey
Juri Belikov, Tallinn University of Technology, Estonia
Maumita Bhattacharya, Charles Sturt University, Australia
Leon Bobrowski, Białystok University of Technology, Poland
Bülent Bolat, Yildiz Technical University, Turkey
Veera Boonjing, King Mongkut's Institute of Technology Ladkrabang, Thailand
Mariusz Boryczka, University of Silesia, Poland
Urszula Boryczka, University of Silesia, Poland
Zouhaier Brahmia, University of Sfax, Tunisia
Stephane Bressan, National University of Singapore
Peter Brida, University of Zilina, Slovakia
Andrej Brodnik, University of Ljubljana, Slovenia
Grażyna Brzykcy, Poznań University of Technology, Poland
Robert Burduk, Wrocław University of Science and Technology, Poland
Aleksander Byrski, AGH University of Science and Technology, Poland

David Camacho, Universidad Autonoma de Madrid, Spain
Tru Cao, Ho Chi Minh City University of Technology, Vietnam
Frantisek Capkovic, Institute of Informatics, Slovak Academy of Sciences, Slovakia
Oscar Castillo, Tijuana Institute of Technology, Mexico
Dariusz Ceglarek, Poznań High School of Banking, Poland
Zenon Chaczko, University of Technology, Sydney
Goutam Chakraborty, Iwate Prefectural University, Japan
Somchai Chatvichienchai, University of Nagasaki, Japan
Chun-Hao Chen, Tamkang University, Taiwan
Rung-Ching Chen, Chaoyang University of Technology, Taiwan
Shyi-Ming Chen, National Taiwan University of Science and Technology
Suphamit Chittayasothorn, King Mongkut's Institute of Technology Ladkrabang, Thailand
Sung-Bae Cho, Yonsei University, South Korea
Kazimierz Choroś, Wrocław University of Science and Technology, Poland
Kun-Ta Chuang, National Cheng Kung University, Taiwan
Dorian Cojocaru, University of Craiova, Romania
Jose Alfredo Ferreira Costa, Federal University of Rio Grande do Norte (UFRN), Brazil
Ireneusz Czarnowski, Gdynia Maritime University, Poland
Theophile Dagba, University of Abomey-Calavi, Benin
Quang A. Dang, Vietnam Academy of Science and Technology, Vietnam
Tien V. Do, Budapest University of Technology and Economics, Hungary
Grzegorz Dobrowolski, AGH University of Science and Technology, Poland
Habiba Drias, University of Science and Technology Houari Boumediene, Algeria
Maciej Drwal, Wrocław University of Science and Technology, Poland
Ewa Dudek-Dyduch, AGH University of Science and Technology, Poland
El-Sayed M. El-Alfy, King Fahd University of Petroleum and Minerals, Saudi Arabia
Keiichi Endo, Ehime University, Japan
Sebastian Ernst, AGH University of Science and Technology, Poland
Nadia Essoussi, University of Carthage, Tunisia
Rim Faiz, University of Carthage, Tunisia
Simon Fong, University of Macau, Macao
Dariusz Frejlichowski, West Pomeranian University of Technology, Szczecin, Poland
Takuya Fujihashi, Ehime University, Japan
Hamido Fujita, Iwate Prefectural University, Japan
Mohamed Gaber, Birmingham City University, UK
Ford Lumban Gaol, BINUS University, Indonesia
Dariusz Gąsior, Wrocław University of Science and Technology, Poland
Janusz Getta, University of Wollongong, Australia
Daniela Gifu, University Alexandru Ioan Cuza of Iasi, Romania
Dejan Gjorgjevikj, Ss. Cyril and Methodius University in Skopje, Macedonia

Barbara Gładysz, Wrocław University of Science and Technology, Poland
Daniela Godoy, ISISTAN Research Institute, Argentina
Ong Sing Goh, Universiti Teknikal Malaysia Melaka, Malaysia
Antonio Gonzalez-Pardo, Universidad Autonoma de Madrid, Spain
Yuichi Goto, Saitama University, Japan
Manuel Grana, University of Basque Country, Spain
Janis Grundspenkis, Riga Technical University, Latvia
Jeonghwan Gwak, Seoul National University, South Korea
Quang-Thuy Ha, VNU-University of Engineering and Technology, Vietnam
Dawit Haile, Addis Ababa University, Ethiopia
Pei-Yi Hao, National Kaohsiung University of Applied Sciences, Taiwan
Marcin Hernes, Wrocław University of Economics, Poland
Francisco Herrera, University of Granada, Spain
Kouichi Hirata, Kyushu Institute of Technology, Japan
Bogumiła Hnatkowska, Wrocław University of Science and Technology, Poland
Huu Hanh Hoang, Hue University, Vietnam
Quang Hoang, Hue University of Sciences, Vietnam
Van-Dung Hoang, Quang Binh University, Vietnam
Jaakko Hollmén, Aalto University, Finland
Tzung-Pei Hong, National Univesity of Kaohsiung, Taiwan
Mong-Fong Horng, National Kaohsiung University of Applied Sciences, Taiwan
Jen-Wei Huang, National Cheng Kung University, Taiwan
Yung-Fa Huang, Chaoyang University of Technology, Taiwan
Maciej Huk, Wrocław University of Science and Technology, Poland
Zbigniew Huzar, Wrocław University of Science and Technology, Poland
Dosam Hwang, Yeungnam University, South Korea
Roliana Ibrahim, Universiti Teknologi Malaysia, Malaysia
Dmitry Ignatov, National Research University Higher School of Economics, Russia
Lazaros Iliadis, Democritus University of Thrace, Greece
Hazra Imran, University of British Columbia, Canada
Agnieszka Indyka-Piasecka, Wrocław University of Science and Technology, Poland
Mirjana Ivanovic, University of Novi Sad, Serbia
Sanjay Jain, National University of Singapore, Singapore
Jarosław Jankowski, West Pomeranian University of Technology, Szczecin, Poland
Joanna Jędrzejowicz, University of Gdańsk, Poland
Piotr Jędrzejowicz, Gdynia Maritime University, Poland
Janusz Jeżewski, Institute of Medical Technology and Equipment ITAM, Poland
Geun Sik Jo, Inha University, South Korea
Kang-Hyun Jo, University of Ulsan, South Korea
Jason Jung, Chung-Ang University, South Korea
Przemysław Juszczuk, University of Economics in Katowice, Poland
Janusz Kacprzyk, Systems Research Institute, Polish Academy of Sciences, Poland
Tomasz Kajdanowicz, Wrocław University of Science and Technology, Poland
Nadjet Kamel, University Ferhat Abbes Setif1, Algeria

Hyun-Deok Kang, Ulsan National Institute of Science and Technology, South Korea
Mehmet Karaata, Kuwait University, Kuwait
Nikola Kasabov, Auckland University of Technology, New Zealand
Arkadiusz Kawa, Poznań University of Economics and Business, Poland
Rafał Kern, Wrocław University of Science and Technology, Poland
Zaheer Khan, University of the West of England, UK
Manish Khare, Dhirubhai Ambani Institute of Information and Communication Technology, India
Chonggun Kim, Yeungnam University, South Korea
Marek Kisiel-Dorohinicki, AGH University of Science and Technology, Poland
Attila Kiss, Eötvös Loránd University, Hungary
Jerzy Klamka, Silesian University of Technology, Poland
Goran Klepac, Raiffeisen Bank, Croatia
Shinya Kobayashi, Ehime University, Japan
Marek Kopel, Wrocław University of Science and Technology, Poland
Raymondus Kosala, BINUS University, Indonesia
Leszek Koszałka, Wrocław University of Science and Technology, Poland
Leszek Kotulski, AGH University of Science and Technology, Poland
Martin Kotyrba, University of Ostrava, Czech Republic
Jan Kozak, University of Economics in Katowice, Poland
Adrianna Kozierkiewicz, Wrocław University of Science and Technology, Poland
Bartosz Krawczyk, Virginia Commonwealth University, USA
Ondrej Krejcar, University of Hradec Kralove, Czech Republic
Dalia Kriksciuniene, Vilnius University, Lithuania
Dariusz Król, Wrocław University of Science and Technology, Poland
Marek Krótkiewicz, Wrocław University of Science and Technology, Poland
Marzena Kryszkiewicz, Warsaw University of Technology, Poland
Adam Krzyzak, Concordia University, Canada
Jan Kubíček, VSB—Technical University of Ostrava, Czech Republic
Tetsuji Kuboyama, Gakushuin University, Japan
Elżbieta Kukla, Wrocław University of Science and Technology, Poland
Julita Kulbacka, Wrocław Medical University, Poland
Marek Kulbacki, Polish-Japanese Academy of Information Technology, Poland
Kazuhiro Kuwabara, Ritsumeikan University, Japan
Halina Kwaśnicka, Wrocław University of Science and Technology, Poland
Annabel Latham, Manchester Metropolitan University, UK
Bac Le, University of Science, VNU-HCM, Vietnam
Hoai An Le Thi, University of Lorraine, France
Yue-Shi Lee, Ming Chuan University, Taiwan
Florin Leon, Gheorghe Asachi Technical University of Iasi, Romania
Horst Lichter, RWTH Aachen University, Germany
Kuo-Sui Lin, Aletheia University, Taiwan
Igor Litvinchev, Nuevo Leon State University, Mexico
Rey-Long Liu, Tzu Chi University, Taiwan

Doina Logofătu, Frankfurt University of Applied Sciences, Germany
Edwin Lughofer, Johannes Kepler University Linz, Austria
Lech Madeyski, Wrocław University of Science and Technology, Poland
Bernadetta Maleszka, Wrocław University of Science and Technology, Poland
Marcin Maleszka, Wrocław University of Science and Technology, Poland
Petra Maresova, University of Hradec Kralove, Czech Republic
Urszula Markowska-Kaczmar, Wrocław University of Science and Technology, Poland
Mustafa Mat Deris, Universiti Tun Hussein Onn Malaysia, Malaysia
Takashi Matsuhisa, Karelia Research Centre, Russian Academy of Science, Russia
Tamás Matuszka, Eötvös Loránd University, Hungary
João Mendes-Moreira, University of Porto, Portugal
Mercedes Merayo, Universidad Complutense de Madrid, Spain
Jacek Mercik, WSB University in Wrocław, Poland
Radosław Michalski, Wrocław University of Science and Technology, Poland
Peter Mikulecky, University of Hradec Kralove, Czech Republic
Marek Miłosz, Lublin University of Technology, Poland
Kazuo Misue, University of Tsukuba, Japan
Jolanta Mizera-Pietraszko, Opole University, Poland
Leo Mrsic, IN2data Ltd Data Science Company, Croatia
Agnieszka Mykowiecka, Institute of Computer Science, Polish Academy of Sciences, Poland
Paweł Myszkowski, Wrocław University of Science and Technology, Poland
Saeid Nahavandi, Deakin University, Australia
Kazumi Nakamatsu, University of Hyogo, Japan
Grzegorz J. Nalepa, AGH University of Science and Technology, Poland
Mahyuddin K. M. Nasution, Universitas Sumatera Utara, Indonesia
Richi Nayak, Queensland University of Technology, Australia
Fulufhelo Nelwamondo, Council for Scientific and Industrial Research, South Africa
Huu-Tuan Nguyen, Vietnam Maritime University, Vietnam
Le Minh Nguyen, Japan Advanced Institute of Science and Technology, Japan
Loan T. T. Nguyen, Nguyen Tat Thanh University, Vietnam
Ngoc-Thanh Nguyen, Wrocław University of Science and Technology, Poland
Quang-Vu Nguyen, Korea-Vietnam Friendship Information Technology College, Vietnam
Thai-Nghe Nguyen, Can Tho University, Vietnam
Van Du Nguyen, Wrocław University of Science and Technology, Poland
Yusuke Nojima, Osaka Prefecture University, Japan
Jerzy Nowacki, Polish-Japanese Academy of Information Technology, Poland
Agnieszka Nowak-Brzezińska, University of Silesia, Poland
Mariusz Nowostawski, Norwegian University of Science and Technology
Alberto Núñez, Universidad Complutense de Madrid, Spain
Manuel Núñez, Universidad Complutense de Madrid, Spain
Richard Jayadi Oentaryo, Singapore Management University, Singapore

Kouzou Ohara, Aoyama Gakuin University, Japan
Tarkko Oksala, Aalto University, Finland
Shingo Otsuka, Kanagawa Institute of Technology, Japan
Marcin Paprzycki, Systems Research Institute, Polish Academy of Sciences, Poland
Rafael Parpinelli, Santa Catarina State University (UDESC), Brazil
Danilo Pelusi, University of Teramo, Italy
Marek Penhaker, VSB—Technical University of Ostrava, Czech Republic
Hoang Pham, Rutgers University, USA
Maciej Piasecki, Wrocław University of Science and Technology, Poland
Bartłomiej Pierański, Poznań University of Economics and Business, Poland
Dariusz Pierzchała, Military University of Technology, Poland
Marcin Pietranik, Wrocław University of Science and Technology, Poland
Elias Pimenidis, University of the West of England, UK
Jaroslav Pokorný, Charles University in Prague, Czech Republic
Nikolaos Polatidis, University of Brighton, UK
Elvira Popescu, University of Craiova, Romania
Piotr Porwik, University of Silesia, Poland
Radu-Emil Precup, Politehnica University of Timisoara, Romania
Andrzej Przybyszewski, Polish-Japanese Academy of Information Technology, Poland
Małgorzata Przybyła-Kasperek, University of Silesia, Poland
Paulo Quaresma, Universidade de Evora, Portugal
Ngoc Quoc Ly, Ho Chi Minh City University of Science, Vietnam
David Ramsey, Wrocław University of Science and Technology, Poland
Mohammad Rashedur Rahman, North South University, Bangladesh
Ewa Ratajczak-Ropel, Gdynia Maritime University, Poland
Leszek Rutkowski, Częstochowa University of Technology, Poland
Tomasz M. Rutkowski, University of Tokyo, Japan
Henryk Rybiński, Warsaw University of Technology, Poland
Alexander Ryjov, Lomonosov Moscow State University, Russia
Keun Ho Ryu, Chungbuk National University, South Korea
Virgilijus Sakalauskas, Vilnius University, Lithuania
Daniel Sanchez, University of Granada, Spain
Rafał Scherer, Częstochowa University of Technology, Poland
Juergen Schmidhuber, Swiss AI Lab IDSIA, Switzerland
Jakub Segen, Polish-Japanese Academy of Information Technology, Poland
Ali Selamat, Universiti Teknologi Malaysia, Malaysia
S. M. N. Arosha Senanayake, Universiti Brunei Darussalam, Brunei Darussalam
Tegjyot Singh Sethi, University of Louisville, USA
Andrzej Siemiński, Wrocław University of Science and Technology, Poland
Dragan Simic, University of Novi Sad, Serbia
Paweł Sitek, Kielce University of Technology, Poland
Krzysztof Ślot, Łódź University of Technology, Poland
Adam Słowik, Koszalin University of Technology, Poland
Vladimir Sobeslav, University of Hradec Kralove, Czech Republic

Kulwadee Somboonviwat, Kasetsart University, Thailand
Kamran Soomro, University of the West of England, UK
Zenon A. Sosnowski, Białystok University of Technology, Poland
Bela Stantic, Griffith University, Australia
Stanimir Stoyanov, University of Plovdiv "Paisii Hilendarski", Bulgaria
Ja-Hwung Su, Cheng Shiu University, Taiwan
Libuse Svobodova, University of Hradec Kralove, Czech Republic
Jerzy Świątek, Wrocław University of Science and Technology, Poland
Andrzej Świerniak, Silesian University of Technology, Poland
Tadeusz Szuba, AGH University of Science and Technology, Poland
Julian Szymański, Gdańsk University of Technology, Poland
Ryszard Tadeusiewicz, AGH University of Science and Technology, Poland
Yasufumi Takama, Tokyo Metropolitan University, Japan
Maryam Tayefeh Mahmoudi, ICT Research Institute, Iran
Zbigniew Telec, Wrocław University of Science and Technology, Poland
Aleksei Tepljakov, Tallinn University of Technology, Estonia
Dilhan Thilakarathne, Vrije Universiteit Amsterdam, Netherlands
Satoshi Tojo, Japan Advanced Institute of Science and Technology, Japan
Krzysztof Tokarz, Silesian University of Technology, Poland
Bogdan Trawiński, Wrocław University of Science and Technology, Poland
Krzysztof Trojanowski, Cardinal Stefan Wyszyński University in Warsaw, Poland
Ualsher Tukeyev, Al-Farabi Kazakh National University, Kazakhstan
Aysegul Ucar, Firat University, Turkey
Olgierd Unold, Wrocław University of Science and Technology, Poland
Joost Vennekens, Katholieke Universiteit Leuven, Belgium
Jørgen Villadsen, Technical University of Denmark, Denmark
Bay Vo, Ho Chi Minh City University of Technology, Vietnam
Eva Volna, University of Ostrava, Czech Republic
Wahyono, Universitas Gadjah Mada, Indonesia
Lipo Wang, Nanyang Technological University, Singapore
Junzo Watada, Waseda University, Japan
Konrad Wojciechowski, Silesian University of Technology, Poland
Krystian Wojtkiewicz, Wrocław University of Science and Technology, Poland
Krzysztof Wróbel, University of Silesia, Poland
Marian Wysocki, Rzeszow University of Technology, Poland
Farouk Yalaoui, University of Technology of Troyes, France
Xin-She Yang, Middlesex University, UK
Tulay Yildirim, Yildiz Technical University, Turkey
Piotr Zabawa, Cracow University of Technology, Poland
Sławomir Zadrożny, Systems Research Institute, Polish Academy of Sciences, Poland
Drago Žagar, University of Osijek, Croatia
Danuta Zakrzewska, Łódź University of Technology, Poland
Constantin-Bala Zamfirescu, Lucian Blaga University of Sibiu, Romania
Katerina Zdravkova, Ss. Cyril and Methodius University in Skopje, Macedonia

Vesna Zeljkovic, Lincoln University, UK
Aleksander Zgrzywa, Wrocław University of Science and Technology, Poland
Qiang Zhang, Dalian University, China
Maciej Zięba, Wrocław University of Science and Technology, Poland
Adam Ziębiński, Silesian University of Technology, Poland
Beata Zielosko, University of Silesia, Poland
Marta Zorrilla, University of Cantabria, Spain

Preface

The ongoing integration of various intelligent information, database systems and technologies—including artificial intelligence, social media, multimedia, cloud computing, and big data processing—leads to important synergy of those fields. It results in new practical solutions and applications as well as broadens the scope of theoretical analyses and approaches. But it also creates previously not considered concerns and problems, related not only to the scale but also to new functionalities of created methods and systems. This includes such issues as reliability and long-term integrity of critical applications, energy efficiency of big data storage, processing, organization, and quality assurance during the development of very complex systems. Continuous research in the fields of intelligent information and database systems provides an opportunity to identify and address those problems by suggesting and creating further solutions and technologies.

The main goal of this book is to present and intensify practical application of current knowledge and state of the art emerging in the field of intelligent information and database systems. Thus, it includes a set of carefully selected research papers presenting new methods, technologies and applications. The represented scope of topics encompasses a wide spectrum of research subjects discussed both from the theoretical and practical points of view. The book covers such fundamental aspects of intelligent information and database systems as artificial and computational intelligence, nature-inspired algorithms and paradigms, collective knowledge and ontologies, computer vision techniques, and multi-dimensional large-scale data processing. At the same time, it presents a number of interdisciplinary topics including social networks, system security, cloud computing, Internet of things, business intelligence, and biomedical data analysis.

This volume in the well-established "Studies in Computational Intelligence" series can be a valuable source of knowledge about current problems and solutions in the field. It contains a selection of 34 chapters based on original research presented as posters during *the Asian Conference on Intelligent Information and Database Systems* (*ACIIDS 2019*) held on 8–11 April 2019 in Yogyakarta, Indonesia. This eleventh edition of the ACIIDS conference was organized by the

Wrocław University of Science and Technology, Poland, together with BINUS University, Indonesia, and their partners.

The book will be a useful resource and reference for researchers and practitioners interested in increasing synergy between artificial intelligence and database technologies as well as for graduate and Ph.D. students in computer science and related fields.

Finally, on behalf of the Steering Committee, Program Committee and Organizing Committee of the ACIIDS 2019 we would like to thank all the authors for sharing their research during the conference. We also extend our thanks to all reviewers for their time, engagement, and expertise. And special gratitude we express to Prof. Janusz Kacprzyk, the editor of this series, and Dr. Thomas Ditzinger from Springer for their interest and support for our project. The help of those people was invaluable in the preparation of this book.

Yogyakarta, Indonesia
April 2019

Maciej Huk
Marcin Maleszka
Edward Szczerbicki

Contents

Comet Assay Classification for Buccal Mucosa's DNA Damage Measurement with Super Tiny Dataset Using Transfer Learning

Afiahayati, Edgar Anarossi, Ryna Dwi Yanuaryska, Fajar Ulin Nuha and Sri Mulyana

Abstract Comet assay or single cell gel electrophoresis assay (SCGE) is a method which is frequently used to measure the damage of DNA. The results of comet assay is a set of comet images, then the comet images are classified to measure the level of DNA damage. Currently, there are several softwares for comet assay image analysis, both free and commercial. Commercial software is quite expensive, while free software is limited, especially for buccal mucosa cell and super tiny comet image dataset. In this research, we propose a classification model for comet assay with super tiny image dataset which is taken from buccal mucosa cells. We propose a transfer learning based convolutional neural network (CNN) model. We have compared the transfer learning model with CNN-support vector machine (SVM) and ordinary CNN. In our experiments, we use super tiny dataset consisting of 73 images. Our transfer learning model gives an accuracy 70.5%, while CNN-SVM gives 62.3% and ordinary CNN gives 63.5%. We also compare our transfer learning model with most frequently used, free comet assay analysis software, OpenComet. Open-Comet gives an accuracy 11.5%. Our transfer learning model is promising for comet assay for buccal mucosa cell and super tiny dataset.

Keywords Transfer learning · Tiny dataset · Comet assay · Buccal mucosa cell

1 Introduction

Comet assay is an excellent method for detecting DNA damage due to its versatility and sensitivity [1]. Comets can be identified and scored by visual inspection by an expert, but this method is very time-consuming and subjective. Therefore, image

Afiahayati (✉) · E. Anarossi · F. U. Nuha · S. Mulyana
Department of Computer Science and Electronics, Faculty of Mathematics and Natural Sciences, Universitas Gadjah Mada, Yogyakarta, Indonesia
e-mail: afia@ugm.ac.id

R. D. Yanuaryska
Department of Radiology Dentomaxillofacial, Faculty of Dentistry, Universitas Gadjah Mada, Yogyakarta, Indonesia

M. Huk et al. (eds.), *Intelligent Information and Database Systems: Recent Developments*, Studies in Computational Intelligence 830,
https://doi.org/10.1007/978-3-030-14132-5_22

analysis software packages were developed [2]. Several tools have been developed to analyze comet assay images. Most of the tools use geometric measuring of the comet's core and tail [3–5]. Currently, there are some softwares for comet assay image analysis, both free and commercial. Commercial software is quite expensive, while free software is limited, especially for buccal mucosa cell. Comet image from different sample sources, for instance mucosa swab and cell line, have slightly different geometric algorithm. It affects the accuracy of the data. Comet images data are hard to obtain, therefore the dataset is limited.

In this research, we propose, an artificial intelligence method specifically developed to score the comet image obtained from buccal mucosa cells. There are two main challenges for this problem domain: (1) comet images obtained from buccal mucosa cells are not like ideal cells since there is a swab activity; (2) the dataset is very tiny. One of artificial intelligence algorithms that could be used to solve image classification problems are Convolutional Neural Network (CNN). CNN is a machine learning method that is designed to process visual information [6–10]. CNN could classify an image based of the patterns learned at the training image sets. The learned patterns then could be used to classify new images. CNN has good performance with a big dataset, while our comet image dataset is super tiny. Therefore, we present a transfer learning based convolutional neural network (CNN) model to classify comet images to measure the level of DNA damage. This work is an extended version of authors' work [11].

2 Data and Methods

2.1 Comet Assay Data

The used comet assay images are cropped image from a microscope slide images of buccal mucosal cells. The total number of images are 73 images. There are 5 levels of the DNA damage. Level 0 represents the lowest damage (the best DNA) while level 4 represents the highest damage (the worst DNA). Those levels are represented as classes in our classification models. Example of comet assay's microscope slide image and samples of comet assay images for each class could be seen in Figs. 1 and 2.

2.2 Pre-processing

In this research, there are 5 stages of pre-processing that produced 2 type of datasets: (1) dataset without data augmentation, and (2) dataset with data augmentation. Pre-processing stages are as follows:

Image Size Normalization Image size normalization uses the Gaussian Pyramid Multi-Scaling Analysis [12] process to expand or reduce the size of image to get as

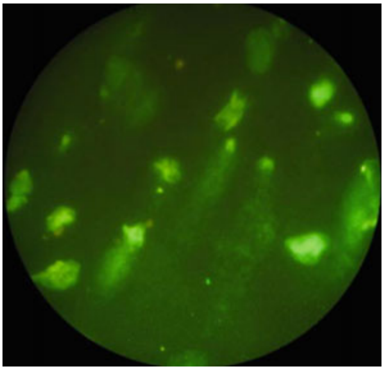


Fig. 1 Comet assay's microscope slide image

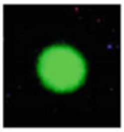
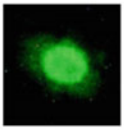
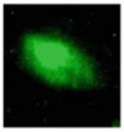
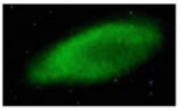
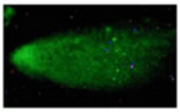
Class	*Comet Assay Image*
0	
1	
2	
3	
4	

Fig. 2 Samples of comet assay images for each class

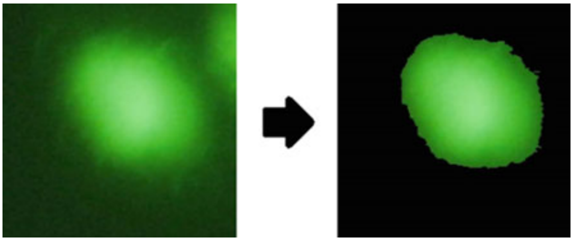


Fig. 3 Comet assay's segmentation

close to 100 × 100 pixel. After that a resize process is done using bilinear interpolation to reach 100 × 100 pixel.

Noise Data Removal Removing images from each class with a similarity percentage to other images in the same class with lower than the predefined similarity threshold. This process is used because the size of dataset are very small, so if the noise images are learned, there will be no other image that could be used to validate if the images are from the respective classes.

Image Segmentation Comet assay objects are separated from the background or other comets inside the image. This process is done to ease the machine learning model at learning the comet assay images. Segmentation is done by using threshold parameter at HSV color space. An example of the segmentation result could be seen at Fig. 3.

RGB Color Space Conversion to Grayscale Color Space Converting the image's RGB color space to Grayscale color space because for the most part of the image, only Green dimension of the RGB has positive values, meanwhile the Red and Blue dimension are zeroes.

Data Augmentation Some of the augmentations used are horizontal and vertical flipping, random cropping, random distortion, and resize. Each classes are augmented to 50 samples per class.

2.3 Classification Models

In order to get good accuracy from such a super dataset, in this research we tried several classification models that could process image data, as follows: (1) Ordinary Convolutional Neural Network, (2) CNN-Support Vector Machine, (3) Transfer Learning (our proposed method).

Ordinary Convolutional Neural Network (CNN) CNN is one of deep learning algorithm variants that is used to process 2 dimensional data. Some parameters that affects a CNN models are convolution layer, pooling layer, activation function, data being used, and how the training is done. Architecture of the CNN model used in this research is represented in Table 1 and Fig. 4.

Table 1 Specification of the CNN model

Layer	Information
Input layer	Image size 100 × 100
Convolution layer	64 filter, size 15 × 15
Pooling layer	Size 2 × 2, stride 2, 2
Convolution layer	256 filter, size 11 × 11
Pooling layer	Size 2 × 2, stride 2, 2
Dense layer	1024 neuron
Dense layer	1024 neuron
Dropout	Dropout rate = 0.2
Output layer	Softmax with 5 neuron

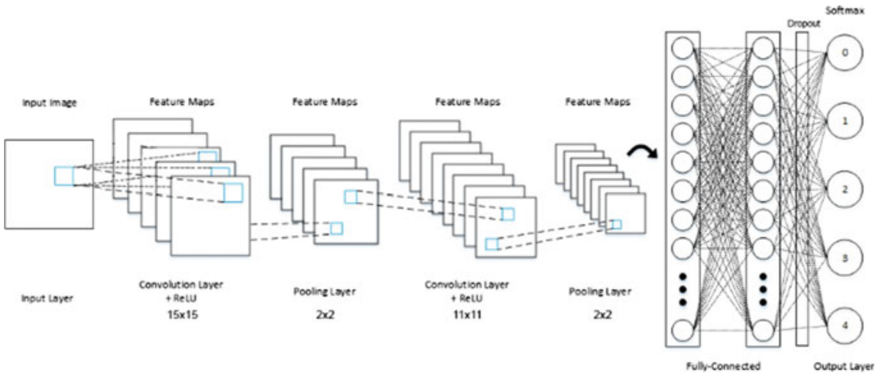


Fig. 4 The architecture of CNN model

The training process is done using the error value calculated from a loss function that is cross entropy and accelerated using ADAGRAD optimizer.

CNN-Support Vector Machine (SVM) SVM solves problems by making N − 1 hyperplane to separate data features with N feature dimensions [13]. Hyperplane or support vector is a line that separates features with as wide margin as possible. Each data will then be categorized into a class according to it's features from the hyperplanes.

Some parameters that affect SVM with RBF as kernel are C and gamma that was tuned in this research. C is a parameter that affects the magnitude of penalty given for each of the misclassified data, meanwhile gamma is a parameter that affects how far the influence of single training example reaches. Best parameters of the SVM model are obtained by using Grid Search. The range of parameters used at grid search is shown in Table 2. The input of SVM in this research is same with the input of the first dense layer at CNN model.

Table 2 Range of parameters used at grid search

Parameter	Range
C	[0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1, 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, 2]
Gamma	[0.0000001, 0.000001, 0.00001, 0.0001, 0.001, 0.01, 0.1, 1]

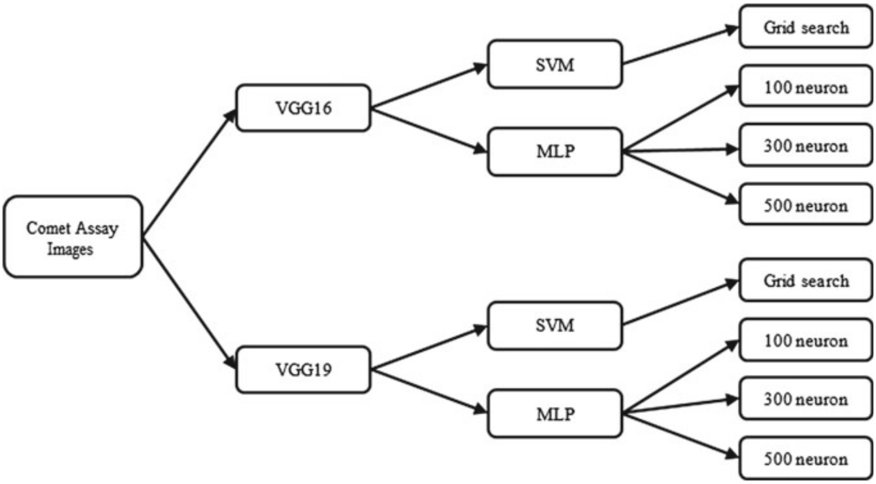


Fig. 5 Classification process using transfer learning

Transfer Learning Transfer learning or inductive transfer is a technique that focused at storing knowledge obtained from solving a problem and applying it to another problem. Transfer learning can be applied at deep learning if a feature learned by a model is general enough to be applied at a target data.

There are two kinds of approach that could be used in transfer learning, that is (1) building a model from scratch and (2) using a pre-trained model. In this research we use a pre-trained model of VGG16 and VGG19 architecture that had been trained on ImageNet dataset which includes 1000 classes [14]. To be able to use the transfer learning model's outputs at classifying comet assay, we used VGG16 and VGG19 outputs as an input at the next classifier that is Multi Layer Perceptron with 1 hidden layer or a Support Vector Machine to get the desired 5 class. Classification process using transfer learning is shown at Fig. 5.

3 Results and Discussion

3.1 Experiment Configuration

For every classification models being tested, 2 types of datasets are used, that is: (1) dataset without data augmentation which includes 61 images, and (2) dataset with data augmentation which includes 250 images. Validations are done using k-fold cross validation with a k value of 4. This research is implemented in Kubuntu 16.10. The CPU is Intel Core i7-6700 CPU @ 3.40 GHz with 16 GB of RAM and only for Ordinary CNN, we use parallel processing with GPU 1080 Ti.

3.2 Results of Ordinary CNN

For the CNN model, testing is exclusively split into 2 parts, which is: (1) testing of parameters related to the quantity of data which includes data augmentation and batch size, and (2) testing of parameters related to the training time which includes learning rate and step size.

The average time of the tests is directly proportional to the size of batch size being tested. Results of data augmentation and batch size tests are shown at Table 3. It is also shown in batch size tests that batch size affects the stability of accuracy and loss function in the training, wherein larger batch size will decrease the amount of change done for each step. From the results obtained from data augmentation and batch size tests at the CNN model, it is shown that unaugmented data has higher accuracy compared to data with augmentations. Furthermore, for the dataset with such a small number, which is $75\% \times 61 = 45$ training data, the best batch size to use in this case is 1 or we could say that the best method for this case is using online learning.

From the results of learning rate and step size experiments, represented at Table 4, it is shown that the best learning rate is 0.00001, whereas using 0.0001 results is a nonoptimal learning while using 0.000001 makes the learning process too slow and

Table 3 Data augmentation and batch size tests results

Batch size	Data	Accuracy	Average time (s)
1	No augmentation	**0.634615406**	237.25
	Augmented	0.500000015	239.5
10	No augmentation	0.519230783	291.25
	Augmented	0.442307696	285
20	No augmentation	0.557692327	291.25
	Augmented	0.519230783	291.25

Table 4 Learning rate and step size tests results

Learning rate	Accuracy (500 steps)	Accuracy (1000 steps)	Accuracy (1500 steps)	Accuracy (2000 steps)
0.0001	0.40385	0.4423	0.40385	0.403846167
0.00001	0.5	0.5577	0.538475	**0.557725**
0.000001	0.365375	0.403825	0.3846	0.403825

Table 5 Best parameter and result of SVM model in CNN-SVM

Data	Accuracy	C	Gamma	Average time (s)
No augmentation	**0.62295082**	1.1	1.00E−07	36
Augmented	0.568	1.5	1.00E−07	345

results in another nonoptimal learning. Too small of a learning rate also results of being stuck in local minima. Size of step size used in this tests isn't too relevant because the step size needed depends on the size of the learning rate.

3.3 Results of CNN-SVM

For the CNN-SVM model, it is shown that most of the tests with unaugmented data results in higher accuracy compared to tests with augmented data. For the gamma parameter, high values result in an optimal model, while small values result in an overfitting model. Furthermore, optimal C value ranges around 0.7–2, so we could say that the penalty given to each misclassification is quite small. The best results of both parameter tests are shown in Table 5.

3.4 Results of Transfer Learning

Transfer Learning model produces the best accuracy compared to the previous models. Similar to previous models, classifications using VGG16 as a feature extractor also results in higher average accuracy when using unaugmented datasets compared to using augmented datasets. From the 4 types of classifiers used to process VGG16's output, it is shown that 1 hidden layer MLP with 300 neuron results in the highest average accuracy. Results of MLP and SVM classification with VGG16 as a feature extractor is shown at Table 6.

Lastly, different from other models' results, Transfer Learning using VGG19 as a feature extractor produces higher accuracy when using augmented datasets rather than unaugmented datasets. This could be caused by 3 extra weight layers that is owned by VGG19 therefore VGG19 could store more features from more data.

Table 6 Results of transfer learning using MLP and SVM as classification methods with VGG16 as a feature extractor

Data/architecture	SVM	MLP 100	MLP 300	MLP 500
No augmentation	0.6202	**0.7049**	**0.7049**	0.639
Augmented	0.6272	0.664	0.68	0.604

Table 7 Results of MLP and SVM classification with VGG19 as a feature extractor

Data/architecture	SVM	MLP 100	MLP 300	MLP 500
No augmentation	0.6238	0.6557	0.623	0.623
Augmented	0.6356	**0.692**	0.668	0.56

Table 8 The accuracy comparison of the overall classification model and OpenComet software [15]

Data/architecture	CNN	CNN-SVM	Transfer learning 1 (VGG16-MLP)	Transfer learning 2 (VGG19-MLP)	OpenComet
No augmentation	0.635	0.623	**0.705**	0.656	0.115
Augmented	0.519	0.568	0.68	0.692	–

From the 4 types of classifiers used to process VGG19's output, it is shown that 1 hidden layer MLP with 100 neuron results in the highest average accuracy. Results of MLP and SVM classification with VGG19 as a feature extractor is shown at Table 7.

The overall model is also compared with the most frequently used, free comet assay analysis software, OpenComet [15]. The data used for OpenComet is same with data for all of our models. The results are represented in Table 8. Our transfer learning 1 model has the highest accuracy compares to other methods. Our transfer learning model gives an accuracy 70.5%, while CNN-SVM gives 62.3% and ordinary CNN gives 63.5% and Open-Comet gives an accuracy 11.5%. CNN model trained with the original comet assay's data gives lower accuracy if compared to transfer learning model that is trained using other dataset (ImageNet), which proves that kernel learning as a feature extractor in Convolutional Neural Network could be done using other dataset with no correlation to the target dataset [16].

4 Conclusion

In this paper, we have classified comet assay images from buccal mucosal cells with super tiny image dataset using several machine learning-based classification models. The CNN model trained with the original comet assay's data gives lower accuracy than the transfer learning model trained with other dataset. Kernel learning as a feature extractor in CNN could use features obtained from learning other dataset at the target dataset, so transfer learning could be used at cases where the target dataset size is too small. Our transfer learning model has the highest accuracy compares to other methods. Our transfer learning model gives an accuracy 70.5%, while CNN-SVM gives 62.3% and ordinary CNN gives 63.5% and Open-Comet gives an accuracy 11.5%. Our transfer learning model is promising for comet assay for buccal mucosa cell and super tiny dataset.

References

1. Collins, A.R., Oscoz, A.A., Giovannelli, L., Brunborg, G., Gaiva, I., Kruszewski, M., Smith, C.C.: The comet assay: topical issues. Mutagenesis **23**, 143–151 (2008)
2. Kumaravel, T.S., Vilhar, B., Faux, S.P., Jha, A.N.: Comet assay measurements: a perspective. Cell Biol. Toxicol. **25**, 53–64 (2009)
3. Sreelatha, G., Rashmi, P., Sathidevi, P.S.: Automatic detection of comets in silver stained comet assay images for DNA damage analysis. In: 2014 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC) on Proceedings, pp. 533–538, IEEE, Guilin, China (2014)
4. Kiziltan, E., Yurtcu, E.: Semi-automatic scoring tool for comet assay. In: 2015 IEEE 15th International Conference on Bioinformatics and Bioengineering (BIBE) on Proceedings, pp. 1–3. IEEE, Belgrade, Serbia (2015)
5. Mondal, N.K., Bhattacharya, P., Ray, M.R.: Assessment of DNA damage by comet assay and fast halo assay in buccal epithelial cells of Indian women chronically exposed to biomass smoke. Int. J. Hyg. Environ. Health **214**, 311–318 (2011)
6. Matsugu, M., Mori, K., Mitari, Y., Kaneda, Y.: Subject independent facial expression recognition with robust face detection using a convolutional neural network. Neural Netw. **16**(5–6), 555–559 (2003)
7. Dewa, C.K., Afiahayati.: Suitable CNN weight initialization and activation function for Javanese vowels classification. Proc. Comput. Sci. **144**, 124–132 (2018)
8. Kurniawan, J., Syahra, S.G.S., Dewa, C.K., Afiahayati.: Traffic congestion detection: learning from CCTV monitoring images using convolutional neural network. Proc. Comput. Sci. **144**, 291–297 (2018)
9. Adam, R., Dewa, C.K., Afiahayati.: Recognizing Arabic letter utterance using convolutional neural network. In: 2017 18th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD), pp. 181–186. IEEE, Kanazawa, Japan (2017)
10. Dewa, C.K.: Javanese vowels sound classification with convolutional neural network. In: 2016 International Seminar on Intelligent Technology and Its Applications (ISITIA), pp. 123–128. IEEE, Indonesia (2016)
11. Anarossi, E.: Klasifikasi Kerusakan DNA pada Comet Assay Menggunakan convolutional neural network. Theses. Universitas Gadjah Mada, Indonesia (2018)
12. Bakkouri, Ibtissam; Afdel, K.: Breast tumor classification based on deep convolutional neural networks. In: 2017 International Conference on Advanced Technologies for Signal and Image Processing (ATSIP) on Proceedings, pp. 1–6. IEEE, Fez, Morocco (2017)
13. Cortes, C., Vapnik, V.: Support-vector networks. Mach. Learn. **20**(3), 273–297 (1995)
14. Simonyan, K., Zisserman, A.: Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv, pp. 1–14 (2014)

15. Gyori, B.M., Venkatachalam, G., Thiagarajan, P.S., Hsu, D., Clement, M.: Redox biology OpenComet: an automated tool for comet assay image analysis. Redox Biol. **2**, 457–465 (2014)
16. Yosinski, J., Clune, J., Bengio, Y., Lipson, H.: How transferable are features in deep neural networks. In: 27th International Conference on Neural Information Processing Systems—Volume 2 on Proceedings, MIT Press, Montreal, Canada, pp. 3320–3328 (2014)