scientific data

Check for updates

OPEN

DATA DESCRIPTOR

FARFUM-RoP, A dataset for computer-aided detection of Retinopathy of Prematurity

Morteza Akbari[1], Hamid-Reza Pourreza[1,2] ✉, Elias Khalili Pour[3] ✉, Afsar Dastjani Farahani[3], Fatemeh Bazvand[3], Nazanin Ebrahimiadib[3], Marjan Imani Fooladi[3] & Fereshteh Ramazani K.[1]

Retinopathy of Prematurity (ROP) is a critical eye disorder affecting premature infants, characterized by abnormal blood vessel development in the retina. Plus Disease, indicating severe ROP progression, plays a pivotal role in diagnosis. Recent advancements in Artificial Intelligence (AI) have shown parity with or surpass human experts in ROP detection, especially Plus Disease. However, the success of AI systems depends on high-quality datasets, emphasizing the need for collaboration and data sharing among researchers. To address this challenge, the paper introduces a new public dataset, FARFUM-RoP (Farabi and Ferdowsi University of Mashhad's ROP dataset), comprising 1533 ROP fundus images from 68 patients, annotated independently by five experienced childhood ophthalmologists as "Normal," "Pre-Plus," or "Plus." Ethical principles and consent were meticulously followed during data collection. The paper presents the dataset structure, patient details, and expert labels.

Background & Summary

Retinopathy of Prematurity (ROP) is a potentially blinding eye disorder that primarily affects premature infants. This condition is characterized by abnormal blood vessel development in the retina of the eye in babies born before 31 weeks of gestation. A critical aspect of ROP is 'Plus Disease', a severe manifestation indicating rapid disease progression. This essay delves into the etiology, pathophysiology, clinical features, management, and implications of ROP and Plus Disease, drawing on a range of scholarly sources. ROP is fundamentally linked to prematurity and low birth weight. The underdevelopment of retinal blood vessels in premature infants is a key factor[1]. After birth, exposure to high levels of oxygen in neonatal care can lead to the cessation of normal retinal vascular growth and subsequent neovascularization[2]. Plus Disease is a marker of disease severity in ROP. It is characterized by dilatation and tortuosity of the posterior retinal blood vessels, indicating a significant increase in disease activity[3]. The diagnosis of ROP is clinical, based on retinal examinations. The International Classification of Retinopathy of Prematurity (ICROP) categorizes ROP based on zone, stage, and the presence of Plus Disease (International Committee for the Classification of Retinopathy of Prematurity, 2005). Plus Disease is identified by the comparison of retinal vascular appearance to a standard photograph[4]. Early detection and treatment are crucial. The management of ROP involves regular eye examinations by a pediatric ophthalmologist. Treatment options include laser therapy, cryotherapy, and anti-VEGF injections, primarily aimed at preventing or slowing the progression of retinal detachment[5]. Untreated, severe ROP can lead to complications such as retinal detachment and blindness. Early treatment, especially in cases of Plus Disease, can significantly improve outcomes. However, long-term follow-up is essential as these infants are at risk of developing other visual impairments[6,7].

The application of Artificial Intelligence (AI) in diagnosing ROP and Plus Disease represents a groundbreaking shift in pediatric ophthalmology. Leveraging advanced AI techniques, such as machine learning and deep learning, has the potential to enhance diagnostic accuracy, efficiency, and accessibility. Recent developments in AI involve creating sophisticated algorithms capable of analyzing retinal images to detect ROP and Plus Disease. These algorithms are trained on extensive datasets, enabling them to identify subtle indicators of the disease with high accuracy[8]. Recent studies have shown that AI can achieve parity with or exceed the diagnostic capabilities

[1]Machine Vision Lab., Faculty of Engineering, Ferdowsi University of Mashhad, Mashhad, 9177948974, Iran. [2]Faculty of Engineering, McMaster University, Hamilton, Ontario, L8S 4L7, Canada. [3]Department of Pediatric Ophthalmology, Farabi Eye Hospital, Tehran University of Medical Sciences, Tehran, 1336616351, Iran. ✉e-mail: hpourreza@um.ac.ir; ekhalilipour@gmail.com

Name of Dataset	Gestation Age (week)	Birth Weight (g)	Image per Patient	Diseased Case Images	Not Diseased Case Images	Number of Labelers	Labels
DIAROP[13]		<2000	10	8090	9711	5	Disease/Not Disease
Zhang and *et al.*[14]				5731	6499		AP-ROP/Regular ROP/Normal
Lei and *et al.*[15]	<36.5	<2000	1~24	11168	11793	5	ROP/Normal
Li and *et al.*[16]				3041	8012	6	ROP/Normal
Mao and *et al.*[17]	26–36	740–3400		981	5180		Normal/Pre-plus/Plus
KIDROP[18]	<34	<2000		89	200	3	No-Plus/Plus
Li and *et al.*[19]	<37	<2000	6	3869	14437	2	Normal/Stages
FARFUM-RoP	<34	<2000	1~41	2085	915	5	Normal/Pre-plus/Plus, Diagnostic, Stage

Table 1. Comparing FARFUM-RoP with different ROP private datasets.

of human experts in ROP. AI's ability to detect Plus Disease, with its complex vascular changes, has been notably enhanced, leading to more reliable and early detection[9].

The advent of AI in the diagnosis of ROP represents a significant leap forward in pediatric ophthalmology. However, the efficacy and accuracy of these AI systems heavily rely on the availability and quality of datasets used in their training and validation. It encourages collaboration between researchers, practitioners, and developers if public datasets exist for this purpose. Sharing data enables the pooling of resources and expertise, leading to more innovative and effective AI solutions. This collaborative environment accelerates the pace of AI development and its subsequent impact on ROP diagnosis. However, it is still a challenging endeavor for researchers that they are unable to precisely evaluate and compare their results with each other. This problem comes from the fact that there is no suitable publicly available dataset for this application. To tackle this problem, some researchers have generated private datasets, which are not shared.

In this paper, we present a public dataset of ROP images, called FARFUM-RoP (standing for Farabi and Ferdowsi University of Mashhad's ROP dataset), for computer-aided detection with research and education purposes. It includes ROP fundus images of 68 patients with 1533 images. Five professional childhood ophthalmologists independently annotated the images to "Normal", "Pre-Plus" and "Plus".

Comparing with other RoP Datasets

Currently, there is no publicly accessible image bank related to ROP (Retinopathy of Prematurity), and therefore FARFUM-RoP is being offered as a public resource. For comparison, here and in Table (1), several ROP-related image banks are presented, all of which are private and the details can only be derived from their respective publications. The only public dataset available in ROP field is the valuable dataset HVDROPDB[10], which is the first public dataset for the segmentation of retinal structure in fundus images of preterm infants.

Methods

Ethical considerations. The present study strictly adhered to the ethical principles outlined in the Declaration of Helsinki, ensuring the welfare and confidentiality of all participants. Approval for the research was obtained from the institutional review board of the Tehran University of Medical Sciences (Ethical approval number: IR.TUMS.FARABIH.REC.1400.073). Written informed consent was meticulously acquired from the parents of each participant, granting explicit permission for their infants to undergo imaging and participate in the study. Also, parental consent for the open publication of the data has been waived by Farabi Hospital Ethics Committee for the stated reasons below. To safeguard privacy, expert technicians anonymize all images before inclusion in the dataset, mitigating any risk of privacy breach.

1. Nature of Data: The retinal images collected do not include any personally identifiable information beyond the images themselves. All data is anonymized to ensure the privacy and confidentiality of the pediatric patients.
2. Usage and Storage: The images are stored in the image repository of Farabi Hospital and are to be used in ophthalmology networks and research projects. These images will not be associated with any identifiable information of the children.
3. Open Publication Waiver: Given the anonymized nature of the data and the potential significant contributions to scientific research and public health, the Ethics Committee has decided to waive the requirement for parental consent specifically for the open publication of this dataset. This decision is based on the understanding that the open publication will not compromise the privacy or safety of the patients involved.
4. Ethical Considerations: The committee acknowledges that open publication of this dataset is essential for the advancement of ophthalmological research. The ethical considerations have been balanced against the potential risks, and it is determined that the benefits of openly publishing this anonymized dataset far outweigh any potential risks.
5. Compliance with Policies: This decision aligns with the ethical standards and policies of Farabi Hospital and relevant international guidelines on human data research.

Image acquisition and selection. A database was created consisting of wide-angle posterior retinal images obtained from various preterm infants in Iran from April 2016 to May 2019. These infants met the published

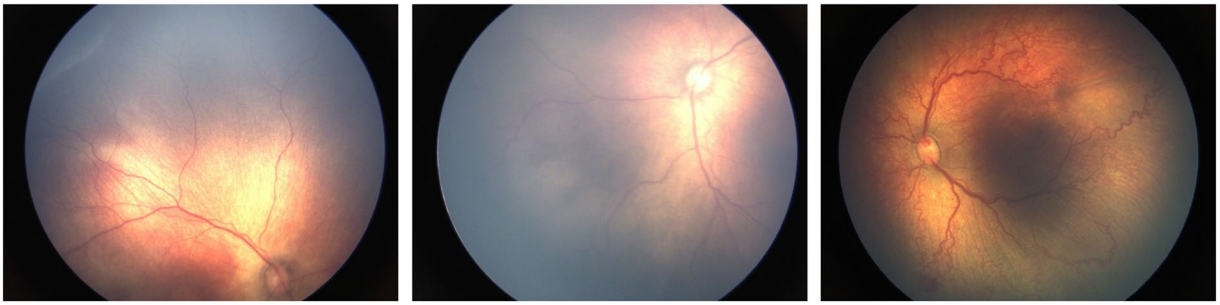

Fig. 1 Examples of Dataset FARFUM-RoP.

Fig. 2 Distribution of patient's images in FARFUM-RoP dataset.

criteria for retinopathy of prematurity (ROP) screening examination, specifically having a birth weight of less than 2000 grams and a gestational age of less than 34 weeks. The images were acquired as part of routine clinical care in the Retinopathy of Prematurity Ward at Farabi Eye Hospital, affiliated with Tehran University of Medical Sciences in Tehran, Iran. Prior to image capture, a mydriatic eye drop, including diluted Phenylephrine and tropicamide drops, was administered to dilate the pupils. The eye drop was administered three times in each eye with a 10-minute interval between each administration. The baby's feeding was discontinued following the third droplet. A total of 2 to 12 pictures were obtained for each eye. RetCam (Clarity Medical Systems, Pleasanton, CA) was used to capture posterior, temporal, superior, inferior, and nasal view pictures of each eye of the newborn.

The challenges faced in collecting retinal images of infants are caused by insufficient patient participation and their limited attention span throughout the image acquisition process. These problems all contribute to the creation of low-quality images with artifacts such as focusing issues, contrast inadequacies, motion blurring, and uneven illumination. The database images are high-quality specialist images that have been selected based on expert opinion from among numerous images taken from the patient. In order to solve image quality concerns, an expert methodically removed about 30 low-quality images from the dataset.

The result, a total bank consisting of 1533 images related to 68 patients was prepared to be labeled by expert doctors in the next step. Some examples of Dataset FARFUM-RoP are shown in Fig. 1.

Dataset

Details. The real names of the patients have been anonymized and encoded as Patient*nn*, where *nn* is a number ranging from 01 to 68. The database consists of 68 compressed.rar files, each named corresponding to the patient's code and containing several.jpg images with dimensions of 1200 by 1600 pixels. For each patient, there are between one to 41 images, averaging 23 images per patient. Figure 2 shows the distribution of patient's images. Additionally, two XLSX files are included. The file named Dataset_ Details.xlsx contains patient information and consists of 68 rows and 4 columns. These columns are: id, which includes codes from Patient01 to Patient68; Patient.BirthWeight, indicating the patient's birth weight in grams; Patient.GestationAge, showing the gestational age at birth in weeks; and finally, Patient.Gender, indicating the patient's gender. Figure 3 shows the distribution of patient's gestation age for our dataset. The other file, Dataset_Labels.xlsx, consists of 1533 rows and eight main columns. These columns are: id, including codes from Patient01 to Patient68; image_name, the names of the image files provided for the patients; five columns for the specialist labels; each of them consists of three sub-columns: retinopathy grade, stage, and diagnostic. Also, the last column presents the collective grade as Label, which specifies the disease label assigned by the five membered Medical Advisory Team. This column contains numbers ranging from 1 to 3, representing the conditions normal, preplus, and plus, respectively.

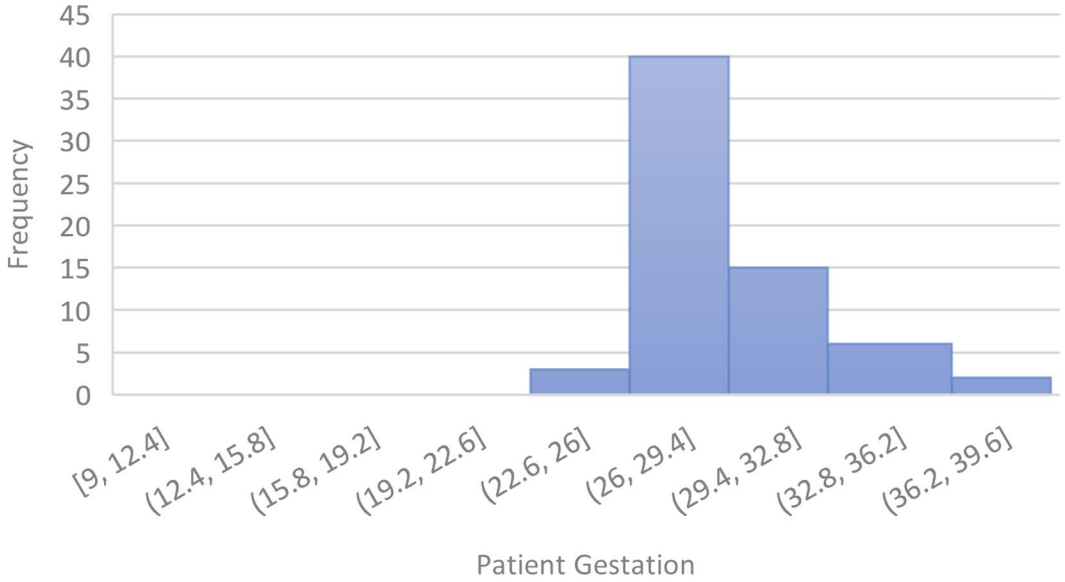

Fig. 3 Distribution of patient's gestation age in FARFUM-RoP dataset.

Data Records

All the data records discussed in this paper are accessible via the Figshare Repository[11], providing an organized format for the research community. The dataset is structured, encompassing 68 individual patients, identified as Patient01 to Patient68.

The patient's dataset is documented in two distinct XLSX files:

1. Dataset Details (Dataset_Details.xlsx): This XLSX file serves as a repository of information regarding the patients and includes birth weight, gestation age, and gender.
2. Dataset Labels Based on Expert Opinions (Dataset_Labels.xlsx): This XLSX file contains labels assigned to each patient based on their ROP fundus images. The images were reviewed by five individual specialists, denoted as A to E, who provided classifications in five different columns:

    a. "Grade" with three conditions: "1 as normal," "2 as preplus," or "3 as plus". "0 means that grade isn't specified".
    b. "Diagnostic" with three conditions: "no treatment," "revisit," or "treatment."
    c. "Stage" with assigned stage values ranging from one to five. "?" means that the ophthalmologist was not sure about the diagnostic.

Additionally, all five specialists, comprising a Medical Advisory Team, collectively reviewed the images and assigned the classifications of "1 as normal," "2 as preplus," and "3 as plus". The resulting consensus was recorded in the "Label" column of this file.

Technical Validation

A software was designed for image labelling, and it has been provided to five retina fellowship-trained experts. These experts are tasked with labelling images from a bank. Among them, one has over 15 years of experience, three have between 10 to 15 years of experience, and the fifth has less than 10 years of experience. All are renowned ophthalmology professors at universities in Iran.

The software features a user interface as shown in Fig. 4. Images are assigned to the specialists, along with relevant information about the patient's weight. For every patient, a set of images, ranging from one to 41 (with an average of 22 images per patient), is curated and prepared. This software allows the specialist to initially review some of a patient's images which are randomly selected. Therefore, the specialist will only be able to see images of the patient that are randomly selected from all the images and can see all these selected images before diagnosing the disease. So, Annotations were performed on individual images, not on entire examinations. After this comprehensive observation, the specialist is then tasked with conducting evaluations across three distinct areas:

1. **Grade Assessment:** Here, the expert must categorize the image into one of three possible conditions: normal, preplus, or plus. This classification is vital in understanding the patient's current ocular state.
2. **Diagnostic Decision:** In this step, the specialist selects one of three potential courses of action based on their analysis: no treatment required, a recommendation for a revisit for further evaluation, or immediate treatment.
3. **Stage Determination:** The expert is tasked with assigning a stage to the condition based on International Committee for the Classification of Retinopathy of Prematurity (ICROP)[12], ranging from one to five. This staging helps in understanding the severity and progression of the ocular issue.

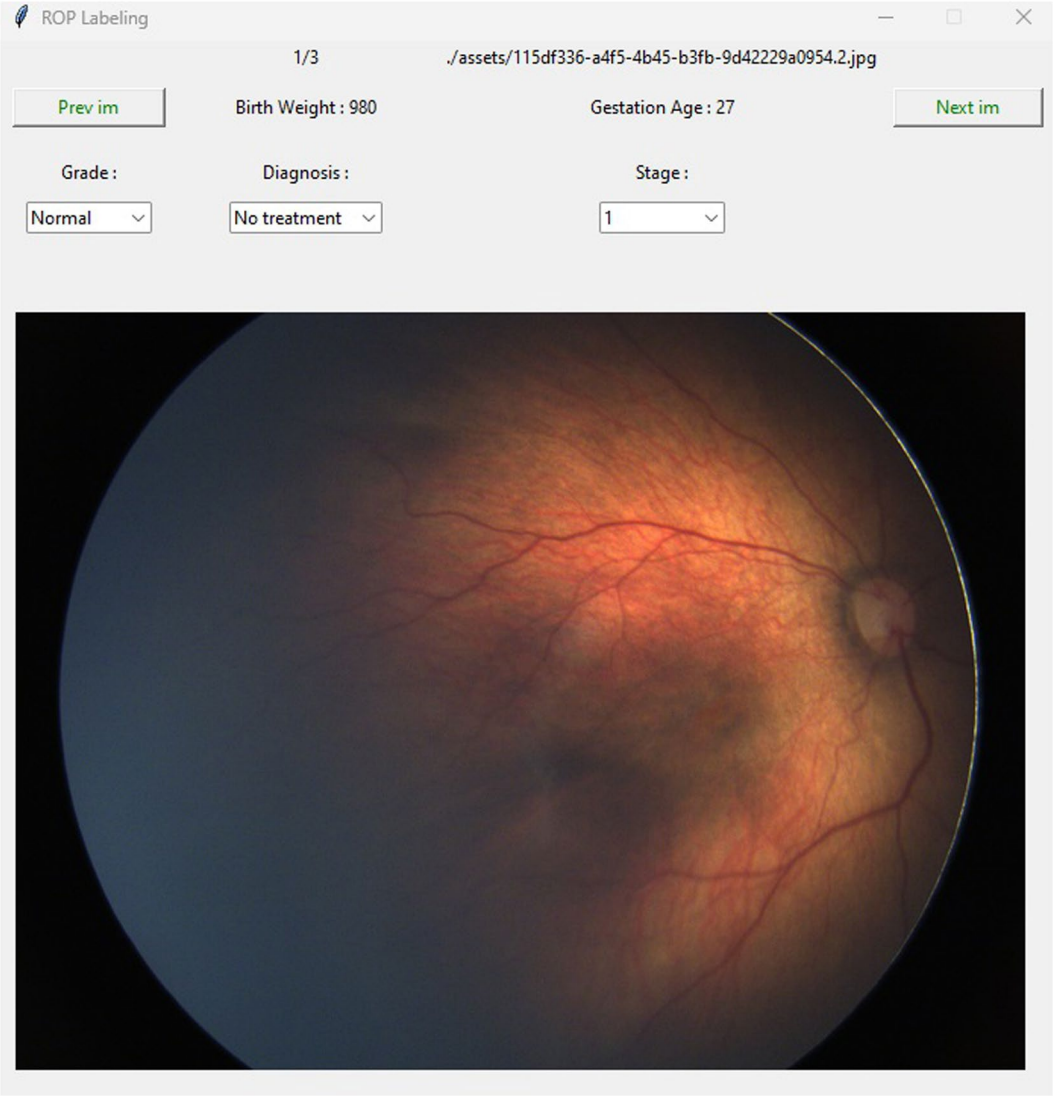

Fig. 4 Labeling software user interface. With the help of this software, the tagger records his idea about the condition, diagnosis, risk of macular edema, and stage of the disease for each patient.

This process ensures a comprehensive evaluation of each image, leveraging the expertise of specialists with varied experiences.

Code availability

For this study no custom code was used.

Received: 9 January 2024; Accepted: 17 September 2024;
Published online: 30 October 2024

References

1. Gilbert, C. Retinopathy of prematurity: A global perspective of the epidemics, population of babies at risk and implications for control. *Early Human Development.* **84**(2), 77–82 (2008).
2. Smith, L. E. & Tasman, W. Retinopathy of prematurity: Current understanding based on new evidence. *Archives of Ophthalmology.* **129**(6), 785–791 (2011).
3. Fierson W. M. Screening examination of premature infants for retinopathy of prematurity. *Pediatrics*. **143**(3) (2019).
4. Wallace, D. K. Plus disease in retinopathy of prematurity: An evolving concept. *Archives of Ophthalmology* **126**(7), 963–964 (2008).
5. Mintz-Hittner, H. A., Kennedy, K. A. & Kennedy, A. Z. Efficacy of Intravitreal Bevacizumab for Stage 3+ Retinopathy of Prematurity. *New England Journal of Medicine* **364**(7), 603–615 (2011).
6. Holmström, G. Long-term follow-up of children with retinopathy of prematurity. *Journal of the American Association for Pediatric Ophthalmology and Strabismus* **11**(5), 454–459 (2007).
7. Shafique M. A. *et al.* Effectiveness of Propranolol in Preventing Severe Retinopathy of Prematurity: A Comprehensive Systematic Review and Meta-Analysis. *American Journal of Ophthalmology*. Nov. 2023.
8. Ting, D. S. W. *et al.* Deep learning in ophthalmology: The technical and clinical considerations. *Progress in Retinal and Eye Research* **81**, 100884 (2021).

9. Brown, J. M. *et al*. Automated diagnosis of plus disease in retinopathy of prematurity using deep learning. *Ophthalmology* **128**(7), 1078–1084 (2021).
10. Agrawal R. *et al*. HVDROPDB datasets for research in retinopathy of prematurity. *Data in Brief*. 1–11 (Nov. 2024).
11. Akbari, M., Pourreza, H. R. & Khalili Pour, E. FARFUM-RoP. *figshare* https://doi.org/10.6084/m9.figshare.c.6721269.v2 (2023).
12. International Committee for the Classification of Retinopathy of Prematurity. The international classification of retinopathy of prematurity revisited. *Archives of Ophthalmology* **123**(7), 991–999 (2005).
13. Attallah, O. DIAROP: automated deep learning-based diagnostic tool for retinopathy of prematurity. *Diagnostics* **11**(11), 2034 (2021).
14. Zhang, R. *et al*. Automatic diagnosis for aggressive posterior retinopathy of prematurity via deep attentive convolutional neural network. *Expert Systems with Applications* (187), 115843.
15. Lei, B. *et al*. Automated detection of retinopathy of prematurity by deep attention network. *Multimedia Tools and Applications* (80), 36341–36360.
16. Li, X. *et al*. Deep multiple instance learning with spatial attention for ROP case classification, instance selection and abnormality localization. Paper presented at: 25th International Conference on Pattern Recognition (ICPR) 2020.
17. Mao, J. *et al*. Automated diagnosis and quantitative analysis of plus disease in retinopathy of prematurity based on deep convolutional neural networks. *Acta ophthalmologica* **98**(3), e339–e345 (2020).
18. Vinekar K. Karnataka internet assisted diagnosis of retinopathy of prematurity. *KIDROP*. Available at: http://kidrop.org/. Accessed December 9, 2023.
19. Li, P. & Liu, J. Early diagnosis and quantitative analysis of stages in retinopathy of prematurity based on deep convolutional neural networks. *Translational Vision Science & Technology*. **11**(5) (2022).

Author contributions

M.A. performed this research as a part of his MSc thesis. H.-R.P. designed the study as the research supervisor. E.K.P., A.D.F., F.B., N.E., M.I.F. and F.R. participated in image analysis and drafted the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41597-024-03897-7.

Correspondence and requests for materials should be addressed to H.-R.P. or E.K.P.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Terms and Conditions

Springer Nature journal content, brought to you courtesy of Springer Nature Customer Service Center GmbH (“Springer Nature”).
Springer Nature supports a reasonable amount of sharing of research papers by authors, subscribers and authorised users (“Users”), for small-scale personal, non-commercial use provided that all copyright, trade and service marks and other proprietary notices are maintained. By accessing, sharing, receiving or otherwise using the Springer Nature journal content you agree to these terms of use (“Terms”). For these purposes, Springer Nature considers academic use (by researchers and students) to be non-commercial.
These Terms are supplementary and will apply in addition to any applicable website terms and conditions, a relevant site licence or a personal subscription. These Terms will prevail over any conflict or ambiguity with regards to the relevant terms, a site licence or a personal subscription (to the extent of the conflict or ambiguity only). For Creative Commons-licensed articles, the terms of the Creative Commons license used will apply.
We collect and use personal data to provide access to the Springer Nature journal content. We may also use these personal data internally within ResearchGate and Springer Nature and as agreed share it, in an anonymised way, for purposes of tracking, analysis and reporting. We will not otherwise disclose your personal data outside the ResearchGate or the Springer Nature group of companies unless we have your permission as detailed in the Privacy Policy.
While Users may use the Springer Nature journal content for small scale, personal non-commercial use, it is important to note that Users may not:

1. use such content for the purpose of providing other users with access on a regular or large scale basis or as a means to circumvent access control;
2. use such content where to do so would be considered a criminal or statutory offence in any jurisdiction, or gives rise to civil liability, or is otherwise unlawful;
3. falsely or misleadingly imply or suggest endorsement, approval , sponsorship, or association unless explicitly agreed to by Springer Nature in writing;
4. use bots or other automated methods to access the content or redirect messages
5. override any security feature or exclusionary protocol; or
6. share the content in order to create substitute for Springer Nature products or services or a systematic database of Springer Nature journal content.

In line with the restriction against commercial use, Springer Nature does not permit the creation of a product or service that creates revenue, royalties, rent or income from our content or its inclusion as part of a paid for service or for other commercial gain. Springer Nature journal content cannot be used for inter-library loans and librarians may not upload Springer Nature journal content on a large scale into their, or any other, institutional repository.
These terms of use are reviewed regularly and may be amended at any time. Springer Nature is not obligated to publish any information or content on this website and may remove it or features or functionality at our sole discretion, at any time with or without notice. Springer Nature may revoke this licence to you at any time and remove access to any copies of the Springer Nature journal content which have been saved.
To the fullest extent permitted by law, Springer Nature makes no warranties, representations or guarantees to Users, either express or implied with respect to the Springer nature journal content and all parties disclaim and waive any implied warranties or warranties imposed by law, including merchantability or fitness for any particular purpose.
Please note that these rights do not automatically extend to content, data or other material published by Springer Nature that may be licensed from third parties.
If you would like to use or distribute our Springer Nature journal content to a wider audience or on a regular basis or in any other manner not expressly permitted by these Terms, please contact Springer Nature at

onlineservice@springernature.com