Рабочая программа дисциплины «Медицинская информатика» (на английском языке) «Medical Informatics»

для образовательной программы «Науки о данных» направления подготовки 01.04.02. Прикладная математика и информатика
уровень - магистр

Разработчик программы
Oleg Pianykh, professor, opiany@gmail.com

Одобрена на заседании департамента анализа данных и искусственного интеллекта «___» __________ 2016 г.
Руководитель департамента анализа данных и искусственного интеллекта Школы
С.О. Куценцов __________

Утверждена Академическим советом образовательной программы «___» __________ 2016 г., № протокола ________________

Академический руководитель образовательной программы
С.О. Куценцов __________________

Москва, 2016

Настоящая программа не может быть использована другими подразделениями университета и другими вузами без разрешения подразделения-разработчика программы.
Medical Informatics

Course Syllabus

I. Introduction: Subject and background

Tutor and author
Professor Oleg S. Pianykh, Professor of National Research University Higher School of Economics, Professor of, PhD

Teaching assistant
?

Summary
Medical Informatics (MI) is a new, exponentially-growing field, where information sciences meet modern clinical applications. The main goal of this class is to introduce HSE students to the broad spectrum of MI problems and applications, and to provide the students with the skills necessary for conducting professional MI work.

Prerequisites
Despite its “medical” content, the class is heavily mathematical, borrowing many math/IT concepts to tackle MI problems. Good programming skills (in any major programming language – Matlab will be used in the class), basic knowledge of statistics, differential/integral equations, function optimization methods, linear algebra would be a must. Optional skills would include knowledge of digital imaging formats (JPEG, bitmap), digital data compression techniques, networking (TCP/IP), real-time programming, cryptography, simulation modeling. All students are advised to become familiar with Matlab programming before taking this class.

Aims
- To develop fundamental knowledge of concepts underlying medical informatics projects
- To develop practical skills needed in modern digital medicine.
- To explain how math and information sciences can contribute to building better healthcare
- To give a hands-on experience with real-world medical data analysis
To develop applied experience with medical software, programming, applications and processes

**Background and outline**

Medical Informatics (MI) is a very new field that bridges math, computer science, and medicine. A fascinating mix of unusual problems and non-trivial solutions that directly impact our lives. Yet – and unfortunately – MI is still not offered by most universities, and the lack of this training is becoming more and more apparent. Today’s graduates have to be more prepared for solving the fundamental MI problems and for doing applied MI work.

This course is aimed at providing our students with a solid MI training, which could boost their careers in contemporary digital healthcare. The course is based on the most recent MI developments, brought to the students from the world-leading hospitals.

While the choice of MI, its problems and projects already defines the novelty of this class, we are trying to do our best to provide our students with the most up-to-date learning experience:

- The class is taught online – convenient to attend and follow. Using the most current teaching software packages, the students can fully interact with the instructor and classmates, share desktops, share applications, record class videos, take online tests and quizzes.

- The students work with real clinical data, solving real clinical problems. That is, unlike more conventional science classes, we prepare our students to solve real-world problems by working on these problems in the class.

- Independent, creative work is emphasized. The class includes several mini-projects, which each student has to design and implement from scratch. Instead of “following the script” we accept the fact that several optimal solutions may be possible in most clinical projects, but only the very best will survive.

- Critical thinking is emphasized. The goal of each class project is to develop a solution that can be used in real life – with noisy data, imperfect practices, human errors, diverse equipment. We teach our students to take clinical data “as is”, and to make most efficient use of what’s available.

- The students get a great chance to learn math and information science in the most applied, “live” way. Although many of students are formally trained in math, they often lack the applied component, connecting any mathematical theory to the underlying reality. In
essence, this class provides a perfect illustration to how math, information and data sciences can help real people in real life, via improved and more efficient healthcare. This also provides the students with a good motivational experience to use their training practically, and to excel in taking practical advantage of their formal knowledge.

This class topic is new to HSE and Russian universities in general – and this is precisely the void we are trying to fill. Modern healthcare needs information science more than anything else, and in return, the information science progress can be driven by the demanding healthcare applications. MI programs start gaining their momentum in leading universities abroad, which is another reason for HSE to cease the opportunity and to offer a competitive class in this field.

**Teaching notes**
The class is offered online, with class material being rather complex and sometimes unusual. Therefore, full student engagement and interaction with the instructor becomes the key to the class’ success. The students should not be taking this course in some “half-duplex”, listening-only mode, which completely defeats the purpose of practical MI training.

To keep the students as engaged as possible, we use a combination of teaching tools and methodology:

- Good online teaching software, blending ease of access with complete interactive experience.
- Constant interaction with the students. In this calls, we remove the line between “lectures” and “practical work”. Instead, the classes are built in a dialog format, when the students are constantly asked questions they have to answer.
- Class projects. While homeworks are meant to demonstrate the understanding of the current class material, we use small class projects (such as building a teleradiology portal) to help students develop their practical MI skills.
- Online “office hours”. Some students may run into problems or fall behind, therefore we provide additional online office hours to help. They are usually scheduled on request, for one or many students, based on their current needs and questions. These “office hours” are usually recorded as well, so that the other students could have full access to their content.
**Teaching outcomes**
The main outcome of this class is to train a student to do practical MI work. Modern hospitals crave for professionals with solid math/information background, but these individuals have to be trained to work with clinical problems. Our goal is to provide this training.
Career-wise, we expect our students to be able to develop into skilled MI researchers or managers (such as CMIO, Chief Medical Information Officer in a hospital).
Finally, from the formal training point of view, we expect our students to become fluent in clinical data acquisition, processing and management, in the areas outlined in the schedule.

**Recommendations to the students**
To the students:
This class is meant to be interesting and fun, and it’s meant to help you unveil a completely new area of human knowledge. But that’s what could contribute to its complexity as well. To anyone thinking about taking this class I would suggest the following:

- Take it only if you are interested in learning something new, and if you are interested in MI in particular
- Be prepared to work
- Be independent, and look for new, unusual solutions. That is, do not simply copy something in existence, do not do “group solutions” – do your own work. If later on you decide to work in a hospital, it will be your work and your responsibility to get things done, so train your own head, don’t pick from the others.
- Do not miss/skip classes and homework. First, homework grades will be responsible for the bulk of your class grade. Second, each class is dedicated to a different area, and you do not want to miss any of them.
- Ask questions. If you are not asking, you are wasting your time.
## II. Schedule

<table>
<thead>
<tr>
<th>No</th>
<th>Topic</th>
<th>Total hours</th>
<th>In class hours</th>
<th>Self-study</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Lectures</td>
<td>Labs</td>
</tr>
<tr>
<td>1</td>
<td>Introduction: What is MI, and what it is not</td>
<td>11</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>Standards: Overview and HL7</td>
<td>11</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>Standards: DICOM</td>
<td>11</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>Making sense of standards</td>
<td>11</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>Computed tomography; Image enhancement</td>
<td>11</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>Computer-Aided Diagnostics (CAD)</td>
<td>11</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>7</td>
<td>Networking and teleradiology (2 lectures)</td>
<td>11</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>8</td>
<td>Security</td>
<td>13</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>9</td>
<td>Scheduling and queuing</td>
<td>13</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>10</td>
<td>Simulation/Modeling in Medicine</td>
<td>13</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>11</td>
<td>Clinical software development; Medical startups (2 lectures)</td>
<td>12</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>12</td>
<td>Medical startups</td>
<td>12</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>13</td>
<td>Unusual applications</td>
<td>12</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td><strong>Total</strong></td>
<td><strong>152</strong></td>
<td><strong>26</strong></td>
<td><strong>38</strong></td>
</tr>
</tbody>
</table>

## III. Assessment

The assessment includes two main components:

- Class homework/projects, assigned after each lecture
- Final exam

The class grade is computed as 80% of homeworks/projects + 20% of the final exam.
In addition to this, student attendance, originality of work and contributions to the class will be taken into account, especially for those balancing between in between two different grades.

**IV. Synopsis**
The course is meant to cover the most principal areas of Medical Informatics (MI). It starts with the introductions into the MI field, its history, and its principal goals. We proceed with the model of modern digital hospital and the two major standards it relies on: DICOM and HL7. Since DICOM provides the foundation to the medical imaging, the class will cover the most critical topics of medical digital image acquisition and analysis: tomography (including Radon transform), image enhancement, CAD, medical image compression. HL7 and its applications will be covered as well, with the emphasis on the clinical data integrity and meaningful use. After this point, the class will evolve into the applications of math/IT aimed at developing optimal clinical processes, including data networking, security, resource management, scheduling, modeling (including pandemic models). The class will end with a series of topics dedicated to clinical application development and implementation (including startups).

**V. Reading**

*Recommended:*

In each class, 2-3 peer-reviewed journal papers will be provided as a required reading on the current class subject.

*Supplementary:*
Oleg S. Pianykh, “Digital Image Quality in Medicine (Understanding Medical Informatics)”, Springer Verlag, 2014
VI. Final exam questions

The final exam will consist of approximately 60 multiple-choice questions, with 30-minute time limit to answer them all. Thus, all questions will be simple but numerous, spanning the entire scope of the course. A few samples are shown below:

1. Which of the following text fragments is HL7?
   a. MSH|^~\&|ADT|N|ADT|MEDSC|200601081527
   b. <html><body><h1>Report</h1><p>PatientName.</p></body></html>
   c. <script type="text/javascript">document.write("<p>"+Date()+"</p>"));
   d. r = input('Enter radius: '); a = pi*r^2;

2. Sinogram for a straight line segment is a
   a. Point
   b. Triangle
   c. Square
   d. Segment

3. Two different algorithms, A and B, were used to remove noise from a diagnostic image. The following are the difference images (Original minus Denoised) for each algorithm. Would you recommend A or B?

   ![Images of A and B]

   a. A
   b. B

4. Recall that to regularize an image, we should minimize the following function, with positive $u_{kn}$:

   $$E(Y) = \sum_n \left[ (x_n - y_n)^2 + \lambda \sum_k u_{kn} (y_k - y_n)^2 \right]$$

   If we replace the original image $X$ by filtered image $Y$ such that $E(Y)$ is minimized when $\lambda = 0.5$, what conclusions can you make about the original and filtered images?

5. One can do tomography without X-Rays:
   a. True
   b. False
6. The following equation defines a bilateral filter:
   \[ c_n = \frac{1 + \sum_{nk} w_{nk} y_n}{x_n + \sum_{N} w_{nk} y_k} \]
   c. True
   d. False

7. Which of the following properties is often enforced with noisy image regularization:
   a. Sharpness
   b. Smoothness
   c. Lambdaess
   d. Brightness

8. It is critical to migrate to the most recent versions of medical standards right after they are released, because:
   a. They provide more functionality
   b. They advance clinical workflow
   c. This is a law
   d. No, it is not critical

9. Temporal data analysis in Radiology becomes possible because
   a. Time stamps are stored in DICOM images
   b. Time stamps are registered by special time-stamping devices
   c. All modalities scan images with the same timing intervals
   d. None of the above

10. What are the two principal types of DICOM data encoding?
    a. VR and non-VR
    b. Private and Standard
    c. Implicit and Explicit
    d. Compressed and Uncompressed

11. Find errors (as many as you see)

| Byte# | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 |
|-------|---|---|---|---|---|---|---|---|---|----|----|----|----|----|----|----|
| Decimal | 16 | 0 | 16 | 0 | P | N | 10 | 0 | S | m | i | t | h | ^ | J | o | e |
| Binary | 10 | 00 | 00 | 10 | 50 | 4E | 0A | 00 | 53 | 6D | 69 | 74 | 68 | 5E | 4A | 6F | 65 |
| g=0010 | e=0010 | VR type | VR length | VR value=Smith^Joe |
| L=10=0x000A |

   a. Even data length
   b. No trailing semicolon
   c. Wrong name separator
   d. VR type must not be present

12. You need to start a teleradiology project, sending CT images to a remote clinic for reports. You measured the network link speed to that clinic; it is 10MB/sec on average. So you consider image compression to send your images faster. The compression software takes 2 seconds to compress (or uncompress) 100 CT images, with average Rcomp=3.2. Is it worth using, and why? CT image size is 0.5 MB.
   a. Yes
   b. No
13. It is critical to migrate to the most recent versions of medical standards right after they are released, because:
   a. They provide more functionality
   b. They advance clinical workflow
   c. This is a law
   d. No, it is not critical

14. Temporal data analysis in Radiology becomes possible because
   a. Time stamps are stored in DICOM images
   b. Time stamps are registered by special time-stamping devices
   c. All modalities scan images with the same timing intervals
   d. None of the above

15. A CT image with “bits allocated” equal to 16 was compressed so that in the compressed format it requires only 2 bits per pixel. What is the compression ratio?
   a. 1
   b. 2
   c. 3
   d. 4

16. Diagnostic image compression ensures that
   a. The original color palette is preserved
   b. The lossy compression error at each pixel is bounded by a chosen threshold
   c. The compression will use only lossless algorithms
   d. The image size remains unchanged

17. The following formula defines
   a. Entropy
   b. Standard Gaussian filter
   c. LoG filter
   d. Denoising filter

\[ E = \frac{-1}{\pi \sigma} \left(1 - \frac{x^2 + y^2}{2\sigma^2}\right) e^{-\frac{x^2 + y^2}{2\sigma^2}} \]

18. Which type of texture analysis will most likely rely on PCA?
   a. Parametric
   b. Non-parametric

19. The following fragment of Matlab code was written to compute bilateral filter for image L at point \((i, j)\):

```matlab
```
for ri = -6:6
    for rj = -6:6
        dL = (L(i, j) - L(i + ri, j + rj))/200; %L1
        w = lambda * exp( dL*dL ) / (1 + rj * rj + ri * ri); %L2 s1
        s1 = s1 + w; %L3
        s2 = s2 + w * L(i + ri, j + rj); %L4
    end
end
end

Which line contains an error?
   a. L1
   b. L2
   c. L3
   d. L4

20. Non-exact conditions in a scheduling program are known as
   a. Hard Constraints
   b. Soft Constraints
   c. Relaxed Constraints
   d. Fixed Constraints

21. What is PHI?
   a. Patient health information
   b. Protected health information
   c. Patient HL7 information
   d. Private health instruction

22. What is HIPAA?
   a. Health Insurance Privacy and Accountability Act
   b. Health Insurance Privacy Access Act
   c. Health Insurance Portability and Accountability Act
   d. HL7 Insurance Portability and Accountability Act

23. Which of the following is/are considered PHI?
   a. Web Universal Resource Locators (URLs)
   b. Certificate / license numbers
   c. Health plan beneficiary numbers
   d. All of the above

24. One needs special DICOM software to find patient name in DICOM files
   a. True
   b. False
25. Can one restore the original data back from the anonymized?
   a. True
   b. False

26. Image watermarking (as discussed in class) is based on:
   a. X-Rays
   b. e-Paper
   c. Anonymization
   d. Steganography

27. Watermarking cannot be destroyed by lossy compression
   a. True
   b. False

VII. Topics for research work
   - Diagnostic image compression
   - Predictive models for patient flow and resource allocation
   - Study of human errors in medical data
   - 3D medical image segmentation

VIII. Typical class projects
   - Codeless teleradiology gateway. This project is illustrated in the appendix below
   - Noise removal in CT images to reduce patient radiation exposure
   - Basic Computer-Aided Diagnostics for liver tumors

The syllabus is prepared by Oleg Pianykh
IX. Appendix: Typical class project – Teleradiology portal

Although we all realize that real clinical projects call for real software development and design, using simple prêt-a-porter software often helps verify the initial concept, and identify the most important processing areas.

This is a freestyle project, and you need to find the easiest solution to the problem below – be creative and use anything you find handy. Here is a good inspiring video: [https://www.youtube.com/watch?v=DVf2JssdYWg](https://www.youtube.com/watch?v=DVf2JssdYWg)

**Problem**: Your hospital wants to implement remote patient consultations. It is assumed that the patients already have their medical data, acquired at the other facilities – on CDs and other digital media. You need to setup a framework, where the patients can upload this data into your hospital’s web site, in the most automated and user-friendly way.

Since text exchange is too easy, we will use images once again. Therefore your solution must have the following parts:

**P1**: Public image upload. Users upload their images into a very intuitive and user-friendly interface, which can be easily run on any public computer.

**P2**: Automated import of newly-loaded images. The newly-uploaded images should be identified as such, and DICOM images must be filtered out for P3.

**P3**: For each medical image identified, your framework should generate some kind of automatic notification (email, tweet, etc) to alarm your hospital staff or doctors.

**Requirements**:

- Parts P2 and P3 must be completely automatic (no manual intervention)
- Each part should be implemented independently, and then integrated into a single process. In other words, you cannot do all three parts in the same application/software
- As little programming as possible. Your goal is to find the existing solutions, not to reinvent the wheel.
“Wow!” – you’ll probably say, – “This sounds like a full-semester project!” Not at all if you look around and assemble the tools already existing around you. Remember, for this assignment, we are looking for a simple “proof of concept” implementation. In other words, you want to use the existing tools to create a simple, yet entirely functional, image upload application.

So let me give you a few hints:

**P1**: Public upload.

- Start from googling “public upload folder”, or think about Dropbox (and similar toys). Or maybe you have a web-hosting account with this functionality already. There are tons of ready (and free) products to solve this problem. Do not write any code for P1.
- You cannot ask your users to FTP or email the images to you. Most of the patients do not have any clue of how FTP works, and their email may not accept large data attachments.

**P2**: Automated import.

- What about some little Matlab code? Once user-uploaded image ends up in your shared folder (Dropbox folder, for instance), you can run a simple Matlab code to scan this folder for DICOM images. This is very much what you had done in our first Matlab assignment, so the only missing part is doing it automatically. But you can write Matlab for-loop, which checks your data folder at each iteration and, once it finds a new file in it, opens it for display, and deletes it from the folder.

Naturally enough, if some of the uploaded files cannot be opened (contain invalid images, for instance), you just ignore them (delete without viewing.). Matlab is an option. If you find another app to do the same job, go ahead, use it. **P3**: Automated notification

Your alternatives may include:

- Sending an email message. For instance, if you know your email server settings, you can check this out: http://www.mathworks.com/support/solutions/en/data/1-3PRRDV/.

- Anything else – you are in charge!

The notification message should include patient name, ID and study date (extracted from the uploaded DICOM file). Also, put your Test ID on the first line.

**Analyzing your project**

Describe your P1 implementation here, outlining its advantages and shortcomings:

Describe your P2 implementation here, outlining its advantages and shortcomings:

Describe your P3 implementation here, outlining its advantages and shortcomings:
Summarize. Now, when you played with your “proof of concept” solution, write three most important recommendations to your development team (going to implement the real, full-scale solution to this problem):

1. 
2. 
3. 

**Important**: this project is not about writing code – it’s about *design and integration*. Finding and using the existing solutions to create a functional image upload application.

**Bonus (optional)**
You can add one more part P4 which does some additional processing of your data – think about anything which would make sense in this project. Once again, be creative in your implementation, describe and explain it below:

---

**Submitting your teleradiology project**

1. If you have a server and can implement your solution there to run 24/7, provide me with a link to the image upload interface here:

   

2. If you do not have a server, I will schedule several time windows to test your solution. During this time, you will need to leave your computer running ready to process my (“patient”) uploads