

# Annotated Plant Pathology Databases for Image-Based Detection and Recognition of Diseases

J. G. A. Barbedo, L. V. Koenigkan, B. A. Halfeld-Vieira, R. V. Costa, K. L. Nechet, C. V. Godoy, M. Lobo Junior, F. R. A. Patrício, V. Talamini, L. G. Chitarra, S. A. S. Oliveira, A. K. N. Ishida, J. M. C. Fernandes, T. T. Santos, F. R. Cavalcanti, D. Terao, F. Angelotti

**Abstract**— Over the last few years, considerable effort has been spent by Embrapa in the construction of a plant disease database representative enough for the development of effective methods for automatic plant disease detection and recognition. In October of 2016, this database, called PDDB, had 2326 images of 171 diseases and other disorders affecting 21 plant species. PDDB size, although considerable, is not enough to allow the use of powerful techniques such as deep learning. In order to increase its size, each image was subdivided according to certain criteria, increasing the number of images to 46,513. Both the original (PDDB) and subdivided (XDB) databases are now being made freely available for academic research purposes, thus supporting new studies and contributing to speed up the advances in the area. Both collections are expected to grow continuously in order to expand their reach. PDDB and XDB can be accessed in the link <https://www.digipathos-rep.cnptia.embrapa.br/>.

**Keywords**— plant pathology, database, deep learning, image processing.

## I. INTRODUÇÃO

**O** RÁPIDO diagnóstico de doenças em plantas de interesse econômico é essencial para assegurar a segurança alimentar e evitar perdas causadas pela

disseminação da patologia. Dois problemas podem prejudicar tal objetivo: 1) o monitoramento permanente de toda a área lavoura por pessoas capazes de detectar as doenças é, na maioria dos casos, inviável; 2) em muitos casos a pessoa que detecta os sintomas não tem o conhecimento necessário para identificar sua causa.

Apesar de haver soluções que exploram a tecnologia como um facilitador para um rápido diagnóstico, normalmente envolvendo uma base de dados de referência a ser consultada pelo usuário [1,2], o surgimento de sistemas totalmente automáticos tem sido lento, especialmente considerando a importância do problema.

A maioria dos métodos propostos até agora são dedicados a identificar uma [3-8] ou algumas poucas doenças [9-11], não possuindo, portanto, a capacidade de lidar com a vasta gama de desordens que podem afetar qualquer espécie de planta. Os poucos métodos que tentam considerar uma gama mais ampla de doenças têm taxas de erros relativamente elevadas [12].

Esta situação não se deve a uma falta de interesse pelos pesquisadores, e sim à falta de uma base de imagens suficientemente representativa para permitir o desenvolvimento de métodos de diagnóstico realmente robustos. Assim, tão importante quanto o desenvolvimento da técnica, é a necessidade de construir bases de dados para treinamento e validação dos métodos propostos.

Nesse contexto, considerável esforço tem sido devotado à construção de uma base de imagens abrangentes. Os desafios envolvidos nessa empreitada são significantes. Há uma grande quantidade de desordens que podem atacar cada espécie de planta, e cada uma dessas desordens produz uma gama de sintomas que dependem de fatores ambientais e contextuais [13].

Como resultado, uma base de imagens verdadeiramente abrangente deveria conter muitos milhares de imagens para cada espécie de plantas. Adicionalmente, a obtenção de imagens de doenças específicas é normalmente uma questão de oportunidade, porque mesmo em experimentos controlados é difícil saber como e se uma dada doença vai se manifestar visualmente.

Portanto, a construção de uma base de dados ampla requer que uma grande quantidade de pessoas trabalhe de modo a alcançar esse objetivo, um horizonte de tempo muito longo ou, provavelmente, ambos. A base de imagens descrita aqui foi construída ao longo de quatro anos e envolveu 17 pesquisadores, muitos dos quais empregaram estudantes e estagiários para ajudar na tarefa. Como resultado, foram coletadas 2326 imagens de 171 doenças e outras desordens acometendo 21 espécies de plantas, as quais foram utilizadas em diferentes estudos [14-17].

Apesar de sua significância, PDDB não é suficientemente

J. G. A. Barbedo, Embrapa Informática Agropecuária, Campinas, SP, Brasil, [jayme.barbedo@embrapa.br](mailto:jayme.barbedo@embrapa.br).

L. V. Koenigkan, Embrapa Informática Agropecuária, Campinas, SP, Brasil, [Luciano.vieira@embrapa.br](mailto:Luciano.vieira@embrapa.br).

B. A. Halfeld-Vieira, Embrapa Meio Ambiente, Jaguariúna, SP, Brasil, [bernardo.halfeld@embrapa.br](mailto:bernardo.halfeld@embrapa.br).

R. V. Costa, Embrapa Milho e Sorgo, Sete Lagoas, MG, Brasil, [rodrigo.veras@embrapa.br](mailto:rodrigo.veras@embrapa.br).

K. L. Nechet, Embrapa Meio Ambiente, Jaguariúna, SP, Brasil, [katia.nechet@embrapa.br](mailto:katia.nechet@embrapa.br).

C. V. Godoy, Embrapa Soja, Londrina, PR, Brasil, [claudia.godoy@embrapa.br](mailto:claudia.godoy@embrapa.br).

M. Lobo Junior, Embrapa Arroz e Feijão, Santo Antônio de Goiás, GO, Brasil, [murillo.lobos@embrapa.br](mailto:murillo.lobos@embrapa.br).

F. R. A. Patrício, Instituto Biológico, Campinas, SP, Brasil, [flavia@biologico.sp.gov.br](mailto:flavia@biologico.sp.gov.br).

V. Talamini, Embrapa Tabuleiros Costeiros, Aracaju, SE, Brasil, [viviane.talamini@embrapa.br](mailto:viviane.talamini@embrapa.br).

L. G. Chitarra, Embrapa Algodão, Campina Grande, PB, Brasil, [luiz.chitarra@embrapa.br](mailto:luiz.chitarra@embrapa.br).

S. A. S. Oliveira, Embrapa Mandioca e Fruticultura, Cruz das Almas, BA, Brasil, [saulo.oliveira@embrapa.br](mailto:saulo.oliveira@embrapa.br).

A. K. N. Ishida, Embrapa Amazônia Oriental, Belém, PA, Brasil, [alessandra.ishida@embrapa.br](mailto:alessandra.ishida@embrapa.br).

J. M. C. Fernandes, Embrapa Trigo, Passo Fundo, RS, Brasil, [mauricio.fernandes@embrapa.br](mailto:mauricio.fernandes@embrapa.br).

T. T. Santos, Embrapa Informática Agropecuária, Campinas, SP, Brasil, [thiago.santos@embrapa.br](mailto:thiago.santos@embrapa.br).

F. R. Cavalcanti, Embrapa Uva e Vinho, Bento Gonçalves, RS, Brasil, [fabio.cavalcanti@embrapa.br](mailto:fabio.cavalcanti@embrapa.br).

D. Terao, Embrapa Meio Ambiente, Jaguariúna, SP, Brasil, [daniel.terao@embrapa.br](mailto:daniel.terao@embrapa.br).

F. Angelotti, Embrapa Semi-Árido, Petrolina, PE, Brasil, [francislene.angelotti@embrapa.br](mailto:francislene.angelotti@embrapa.br).

representativa para permitir o uso de técnicas poderosas tais como “Deep Learning”, as quais tornam possível construir conceitos complexos a partir de conceitos mais simples [18]. Contudo, em muitos casos as imagens contêm uma riqueza de informações na forma de múltiplas lesões e sintomas, cada qual com suas características particulares.

Assim, a fim de aumentar o tamanho da base de dados, as imagens foram divididas em imagens menores contendo lesões individuais ou regiões de sintomas, de acordo com alguns critérios descritos na próxima seção. Como resultado, esta base de imagens expandida (XDB) atualmente contém 46.513 imagens.

PDDB e XDB ainda têm lacunas a serem preenchidas, em termos tanto das desordens contempladas quanto da gama de manifestações visuais representadas na base de imagens. O principal objetivo de tornar essas bases de imagens disponíveis é ajudar outros pesquisadores a conduzir seus estudos e trazer avanços à área de detecção e identificação automática de patologias em plantas. Porém, pretende-se também encorajar outras pessoas a se juntar aos esforços através da submissão de novo conteúdo à base de dados, a qual se espera que continue a crescer continuamente no futuro.

## II. MATERIAIS E MÉTODOS

### A base de imagens PDDB

As imagens na base de imagens foram capturadas usando vários sensores diferentes (*smartphones*, câmeras compactas, câmeras DSLR), e as resoluções variaram de 1 a 24 MPixels. O esquema de classificação primário adotado para a PDDB considera espécies de plantas e desordens como as duas camadas hierárquicas. A Tabela I especifica o número de imagens para cada par planta/desordem.

A maior parte das desordens representadas nas imagens é relacionada a fungos (77%), enquanto 8% são devidas a vírus, 6% a pragas, 3% a bactérias, 2% a fitotoxidez, 2% a algas, 1% a deficiências nutricionais e 1% a senescência. A descrição detalhada das causas das doenças pode ser encontrada em [12].

PDDB inclui 715 imagens capturadas em campo e 1611 imagens capturadas sob condições controladas. Ambos os tipos de imagens são importantes: enquanto um *software* para diagnóstico de doenças normalmente requer validação com imagens capturadas sob condições reais, imagens capturadas sob condições controladas são essenciais nos estágios iniciais de desenvolvimento do algoritmo, uma vez que suas características são mais previsíveis. Isto levou à criação de um dispositivo composto de duas partes: uma caixa de papelão funcionando como uma câmara escura (Fig. 1), e um cartão contendo padrões coloridos e branco-e-preto para correções geométricas e de cor (Fig. 2).

A caixa foi criada tendo *smartphones* em mente, porém ela pode ser usada em conjunto com câmeras compactas, e seu interior é branco a fim de maximizar o espalhamento da luz após o disparo do *flash*. Pouco mais de metade das imagens capturadas em condições controladas foram obtidas usando esse dispositivo.

Apesar da maioria das imagens contem folhas como elemento principal, algumas delas focam em outras partes das plantas. A Tabela II detalha a composição da base de dados de

acordo com a parte da planta que é mostrada na imagem. É importante notar que algumas imagens incluem múltiplas partes da planta, em cujo caso apenas o elemento principal é considerado nessa categorização.

### Crítérios para expansão da base de imagens

Apenas as imagens contendo folhas de plantas foram consideradas na criação da XDB, porque os sintomas visíveis em outras partes das plantas geralmente não são divisíveis. Os fundos de todas as imagens foram removidos manualmente antes da subdivisão.

Cinco tipos diferentes de sintomas foram identificados, cada qual requerendo diferentes critérios de subdivisão, sempre tendo diversidade de características como principal meta. Cada doença pode gerar diferentes tipos de sintomas dependendo de sua severidade e estágio de desenvolvimento. É importante destacar que, para todas as novas imagens resultantes da subdivisão, o tecido saudável ocupou ao menos 20% da área total, para garantir contraste com o tecido doente.

O primeiro tipo de sintoma, pequeno espalhado, consiste de numerosas pequenas lesões ou manchas espalhadas pela superfície da folha (Fig. 3a). Dois critérios foram usados: sintomas relativamente isolados foram tomados individualmente (Fig. 3b), e lesões que faziam parte de aglomerados foram tomadas como um grupo (Fig. 3c).

O segundo tipo de sintoma, grande espalhado, consiste de várias lesões ou manchas grandes espalhadas sobre a superfície da folha (Fig. 4a). Os critérios adotados aqui foram os mesmos do caso anterior, mas devido ao tamanho maior dos sintomas, a caixa delimitando a área ao redor da lesão pode incluir partes de outras lesões, mesmo se estas são relativamente isoladas. Quando isto ocorreu, em cerca de metade dos casos as lesões espúrias foram removidas da imagem (Fig. 4b), e na outra metade as lesões espúrias foram mantidas sem modificações (Fig. 4c). Isto foi feito para aumentar a diversidade de condições.

O terceiro tipo de sintoma, isolado, consiste de lesões ou manchas únicas (Fig. 5a). Nesse caso, como há apenas uma região de interesse, apenas uma nova imagem é gerada a partir da original (Fig. 5b). As únicas exceções a esta regra surgiram em casos em que as lesões continham regiões muito distintas (Fig. 5c).

O quarto tipo de sintoma, generalizado, consiste de grandes lesões espalhadas por toda a folha (Fig. 6a). Devido à ampla variedade de características encontradas neste grupo, os critérios para subdivisão foram definidos de maneira vaga. Primeiro, toda a imagem original (com o fundo removido) é considerada como uma sub-imagem. As subdivisões remanescentes foram realizadas através da identificação de regiões doentes relativamente homogêneas (Fig. 6b e 6c).

Finalmente, o quinto tipo de sintoma, pulverulento, consiste de manchas na superfície da folha causadas pelos esporos de fungos, os quais têm uma aparência de pó (Fig. 7a). Essas manchas normalmente começam separadas, porém elas podem se fundir à medida que a doença evolui, podendo encobrir toda a superfície da folha. Quando as manchas estão isoladas, cada uma delas gera uma nova imagem (Fig. 7b). Quando a doença está mais avançada, a divisão segue as mesmas regras básicas das lesões generalizadas (Fig. 7c).

É interessante notar que as imagens resultantes podem sofrer modificações adicionais (rotação, espelhamento, etc.) a

fim de gerar ainda mais amostras [19].

Espécie (Nome Comum)	Doença (Nome Comum)	Espécie Causal (Nome Científico)	Número de Imagens		Espécie (Nome Comum)	Doença (Nome Comum)	Espécie Causal (Nome Científico)	Número de Imagens		
			PDDB	XDB				PDDB	XDB	
<i>Piper nigrum</i> (Pimenta do Reino)	Antracnose	<i>Colletotrichum gloeosporioides</i>	1	1	<i>Cocos nucifera</i> (Coqueiro)	—	<i>Aceria guerreronis</i>	33	-	
	Podridão Raízes	<i>Phytophthora capsici</i>	5	-		—	<i>Dreschlera incurvata</i>	2	2	
	Queima Fio	<i>Marasmiellus scandens</i>	3	-		—	<i>Stenotarsonemus furcatus</i>	2	-	
	Murcha Amarela Saudável	<i>Fusarium oxysporum</i>	3	-		Lagarta	—	—	4	-
	—	—	—	3		-	Queima folhas	<i>Cytospora palmarum</i>	88	-
<i>Phaseolus vulgaris</i> (Feijão)	Antracnose	<i>Colletotrichum lindemuthianum</i>	21	561	Cochonilha	<i>Aspidiotus destructor</i>	5	583		
	Cercosporiose	<i>Cercospora sp.</i>	2	38	Fumagina	<i>Capnodium spp.</i>	2	-		
	Crestam. bacteriano	<i>Xanthomonas axonopodis</i>	25	118	Lixa grande	<i>Coccostromopsis palmicola</i>	51	609		
	Ferrugem	<i>Uromyces appendiculatus</i>	2	420	Lixa pequena	<i>Phyllachora torrendiella</i>	53	195		
	—	<i>Hedylepta indicata</i>	5	100	M. <i>Cylindrocladium</i>	<i>Cylindrocladium pteridis</i>	5	100		
	Mancha alvo	<i>Corynespora cassicola</i>	24	600	Resinose	<i>Thielaviopsis paradoxa</i>	25	-		
	Mancha bacteriana	<i>Xanthomonas axonopodis</i>	2	38	Mosca branca	<i>Aleurodicus destructor</i>	6	2		
	Mosaico angular	Cowpea mild mottle virus	6	78	Podridão Lasiopodia	<i>Lasiopodia theobromae</i>	2	-		
	Podridão cinzenta	<i>Macrophomina phaseolina</i>	1	-	Lixas grande e pequena	—	21	-		
	Mela	<i>Thanatephorus cucumeris</i>	7	75	Lixa pequena + queima	—	7	-		
<i>Anacardium occidentale</i> (Cajuzeiro)	Oídio	<i>Erysiphe polygoni</i>	12	183	Fitotoxidez (pestic.)	—	2	-		
	Mosaico dourado Saudável	<i>Begomovirus sp.</i>	12	12	Bicho mineiro	<i>Leucoptera coffeella</i>	12	52		
	—	—	2	2	Cercosporiose	<i>Cercospora coffeicola</i>	55	88		
	Fitotoxidez (pestic.)	—	8	939	Cercosp. + ferrugem	—	3	-		
	Alga	—	3	111	<i>Coffea arabica</i> (Cafê)	Ferrugem	<i>Hemileia vastatrix</i>	44	383	
	Mancha angular	<i>Septoria anacardii</i>	8	1332	Mancha aureolada	<i>Pseudomonas syringae</i>	37	1335		
	Antracnose	<i>Colletotrichum gloeosporioides</i>	37	847	Mancha mantegosa	<i>Colletotrichum gloeosporioides</i>	8	175		
	Mofa preto	<i>Pilgeriella anacardii</i>	33	2114	Mancha phoma	<i>Phoma costaricensis</i>	25	52		
	Resinose	<i>Lasiopodia theobromae</i>	36	42	Rizoctoniose	<i>Rhizoctonia solani</i>	2	-		
	Oídio	<i>Oidium anacardii</i>	6	65	Saudável	—	4	-		
<i>Manihot esculenta</i> (Mandioca)	Seca dos ponteiros Larvas	<i>Sclerotium rolfsii</i>	1	-	Antracnose	<i>Elsinoë ampelina</i>	1	-		
	—	—	3	98	Cancro bacteriano	<i>Xanthomonas campestris</i>	13	389		
	Desconhecido	—	4	-	Ferrugem	<i>Phakopsora euvitis</i>	8	458		
	Antracnose	<i>Colletotrichum gloeosporioides</i>	1	-	<i>Vitis vinifera</i> (Videira)	Mancha folhas	<i>Mycosphaera personata</i>	1	32	
	Bacteriose	<i>Xanthomonas axonopodis</i>	18	650	Mildio	<i>Plasmopara viticola</i>	22	173		
	Mosaico nervuras	Vírus mosaico nervuras	17	11	Oídio	<i>Erysiphe necator</i>	29	51		
	Mosaico comum	Vírus mosaico comum	1	1	Degenerescência	Grapevine fanleaf virus	7	31		
	Oídio	<i>Erysiphe polygoni</i>	1	7	Saudável	—	2	-		
	Mancha parda	<i>Gloeosporium fructigenum</i>	2	4	<i>Gossypium hirsutum</i> (Algodão)	Mela	<i>Rhizoctonia solani</i>	32	166	
	Ácaro	<i>Mononychellus tanajoa</i>	10	130	Mancha mirotécio	<i>Myrothecium roridum</i>	27	146		
<i>Citrus spp</i> (Citrus)	Queima folhas	<i>Cercospora vicosae</i>	1	3	Fusariose	<i>Fusarium oxysporum Schlechtend</i>	29	-		
	Mosca galhas	<i>Jatrophobia brasiliensis</i>	2	4	Ramulária	<i>Ramularia areola</i>	36	1712		
	Podridão raízes	<i>Phytophthora sp</i>	58	-	Antracnose	<i>Colletotrichum truncatum</i>	2	-		
	Saudável	—	14	-	Crestamento bacteriano	<i>Pseudomonas savastanoi</i>	56	3794		
	M. branca e parda	—	9	115	Crestamento foliar	<i>Cercospora kikuchii</i>	5	10		
	Alga	—	5	244	Falso carvão	<i>Macrophomina phaseolina</i>	12	-		
	Mancha alternaria	<i>Alternaria alternata f. sp. citri</i>	2	2	Ferrugem	<i>Phakopsora pachyrhizie</i>	65	2267		
	Mancha preta	<i>Guignardia citricarpa</i>	4	-	Fitotoxidez (pestic.)	—	23	2112		
	Cancro	<i>Xanthomonas axonopodis</i>	9	227	Folha carijó	Soybean Mosaic Virus	22	303		
	Clorose	<i>Xylella fastidiosa</i>	27	308	Mancha alvo	<i>Corynespora cassicola</i>	62	1015		
<i>Citrus spp</i> (Citrus)	Feltro	<i>Septobasidium pseudopedicellatum</i>	2	-	<i>Glycine max</i> (Soja)	Mancha mirotécio	<i>Myrothecium roridum Tode</i>	2	8	
	Fumagina	<i>Capnodium spp.</i>	4	148	Podridão Phytophthora	<i>Phytophthora sojae</i>	3	-		
	Leprose	<i>Citrus leprosis virus (CiLV)</i>	18	65	Mildio	<i>Peronospora manshurica</i>	52	2514		
	Mancha aureolada	<i>Pelicularia filamentosa</i>	5	7	Oídio	<i>Microsphaera diffusa</i>	77	222		
	Mancha graxa	<i>Mycosphaera citri</i>	8	271	Septoriose	<i>Septoria glycines</i>	21	439		
	Mosaico amarelo	—	15	440	Mela	<i>Rhizoctonia solani</i>	2	-		
	Verrugose	<i>Elsinoë fawcettii</i>	2	153	Murcha-de-esclerócio	<i>Sclerotium rolfsii</i>	4	-		
	Mofa verde	<i>Penicillium digitatum</i>	10	-	Oídio + ferrugem	—	1	-		
	Queima fio	<i>Marasmiellus scandens</i>	3	-	Ferrugem + m. alvo	—	32	-		
	Podridão floral	<i>Colletotrichum acutatum</i>	2	-	Saudável	—	9	-		
<i>Zea mays</i> (Corn)	Defic. nutricional Saudável	—	11	-	Desconhecido	—	41	873		
	—	—	8	-	<i>Brassica oleracea</i> (Couve)	Oídio	<i>Erysiphe polygoni</i>	3	63	
	Antracnose	<i>Colletotrichum graminicola</i>	7	26	Mancha alternária	<i>Alternaria brassicicola</i>	4	133		
	Enfezamento	<i>Dalbulus maidis</i>	6	5	Saudável	—	1	-		
	Ferrugem branca	<i>Physopella zea</i>	14	889	<i>Cucumis melo L.</i> (Meloeiro)	Mancha aquosa	Acidovorax avenae	2	-	
	Ferrugem polissora	<i>Puccinia polysora</i>	15	3051	Mosaico	Melon mosaic virus	1	1		
	Lixa	<i>Gibberella zea</i>	3	723	Oídio	<i>Podospaera xanthii</i>	5	148		
	Mancha bipolaris	<i>Bipolaris maydis</i>	44	3773	Nematoide	<i>Meloidogyne incognita</i>	1	-		
	Mancha branca	<i>Phaeosphaeria maydis f. imp. Phyllosticta sp</i>	31	779	<i>Theobroma grandiflorum</i> (Cupuaçu)	Podridão Lasiopodia	<i>Lasiopodia theobromae</i>	1	-	
	Mancha diplodia	<i>Stenocarpella maydis</i>	7	18	Vassoura de bruxa	<i>Crinipellis pernicioso</i>	4	-		
Mancha physoderma	<i>Physoderma maydis</i>	8	1072	<i>Oryza sativa</i> (Rice)	Escaldadura	<i>Microdochium oryzae</i>	2	2		
Mancha turcicum	<i>Exserohilum turcicum</i>	46	110	Brusone	<i>Magnaporthe grisea</i>	4	247			

TABELA I

## COMPOSIÇÃO DE PDDB E XDB EM TERMOS DAS DOENÇAS E ESPÉCIES DE PLANTAS.

Espécie (Nome Comum)	Doença (Nome Comum)	Espécie Causal (Nome Científico)	Número de Imagens		Espécie (Nome Comum)	Doença (Nome Comum)	Espécie Causal (Nome Científico)	Número de Imagens	
			PDDB	XDB				PDDB	XDB
<i>Elaeis guineensis</i> (Palma de Óleo)	Antracnose	<i>Botryodiplodia palmarum</i>	1	1	<i>Passiflora edulis</i> (Maracujá)	Antracnose	<i>Colletotrichum gloeosporioides</i>	9	4
	Míldio negro	—	1	30		Cercosporiose	<i>Cercospora sp.</i>	4	95
	Mancha curvulária	<i>Curvularia oryzae</i>	1	-		Cladosporiose	<i>Cladosporium herbarum</i>	19	1
	Fusariose	<i>Fusarium oxysporum</i>	3	-		Mancha bacteriana	<i>Xanthomonas axonopodis</i>	38	169
	Ilhas necróticas	<i>Cercospora elaeidis</i>	1	-		Septoriose	<i>Septoria passiflorae</i>	7	16
	Anel vermelho	<i>Bursaphelenchus cocophilus</i>	4	-		Podridão	<i>Phytophthora cinnamomi</i>	3	-
	Podridão Thielaviopsis	<i>Thielaviopsis paradoxa</i>	4	-		Mancha alternária	<i>Alternaria spp.</i>	1	-
Desconhecido	—	1	-	—	<i>Dione juno juno</i>	6	-		
<i>Saccharum officinarum</i> (Cana de açúcar)	Ferrugem	<i>Puccinia recondata</i>	18	1014	Fusariose	<i>Fusarium solani</i>	1	-	
	Mancha anelar	<i>Leptosphaeria sacchari</i>	43	1657	Mosaico	Mosaic Virus	42	23	
	M. anelar + p. vermelha	—	4	-	Senescência	—	6	6	
	Podridão vermelha	<i>Acidovorax avenae</i> subsp. <i>avenae</i>	49	104	Saudável	—	6	-	
	M. anelar + mosca branca	—	4	-					
Saudável	—	5	-	<i>Carica papaya</i> (Mamoeiro)	Antracnose	<i>Colletotrichum brevisporum</i>	8	-	
<i>Triticum aestivum</i> (Trigo)	Brusone	<i>Magnaporthe oryzae</i>	10	198	Mosaico	<i>Mosaic virus</i>	11	11	
	Ferrugem	<i>Puccinia triticina</i>	13	334	Variola	<i>Asperisporium caricae</i>	7	785	
	Mancha amarela	<i>Drechslera tritici-repentis</i>	2	36	<i>Ananas comosus</i> (Abacaxi)	Broca	<i>Strymon megarus</i>	3	-
	Oídio	<i>Blumeria graminis</i>	19	134	Fusariose	<i>Fusarium guttiforme</i>	10	-	
Saudável	—	7	-	Podridão	<i>Ceratocystis paradoxa</i> Moreau	6	-		
					Total		2326	46513	

TABELA I (CONT.)  
COMPOSIÇÃO DE PDDB E XDB EM TERMOS DAS DOENÇAS E ESPÉCIES DE PLANTAS.TABELA II  
COMPOSIÇÃO DA PDDB EM TERMOS DA PARTE DA PLANTA.

Espécie de Planta	Folha	Flor	Fruta/ Espiga	Caule	Raiz
Pimenta do reino	6	0	2	7	0
Feijão	119	0	9	1	0
Cajuzeiro	91	0	5	34	1
Mandioca	96	0	0	38	0
Citrus	107	2	21	5	0
Coqueiro	244	0	40	24	0
Coffee	177	0	7	6	0
Corn	181	0	0	0	0
Cotton	95	0	0	29	0
Cupuaçu	0	0	3	2	0
Grapevines	58	0	6	19	0
Kale	8	0	0	0	0
Melon	6	0	2	0	1
Oil Palm	7	0	0	9	0
Papaya	18	0	8	0	0
Passion Fruit	103	0	31	6	2
Pinneapple	3	0	11	2	3
Rice	6	0	0	0	0
Soybean	471	0	10	10	0
Sugarcane	123	0	0	0	0
Wheat	51	0	0	0	0
<b>Total</b>	<b>1970</b>	<b>2</b>	<b>155</b>	<b>192</b>	<b>7</b>

## III. DISCUSSÃO

A última coluna da Tabela I mostra a composição da base XDB, resultante da subdivisão das imagens em PDDB. O número de imagens na XDB é significativamente maior para a maioria das doenças, com algumas poucas exceções associadas à presença significativa de imagens de outras partes das plantas que não as folhas. Esse aumento foi maior para doenças que se manifestam na forma de lesões ou manchas numerosas (e.g.

ferrugem polissora em milho) do que para aquelas que geram apenas algumas poucas lesões (e.g. cercosporiose do café).

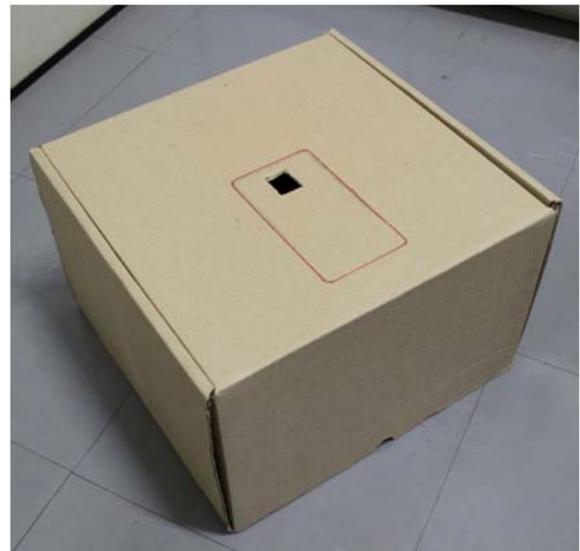


Figura 1. Caixa usada na captura das imagens.

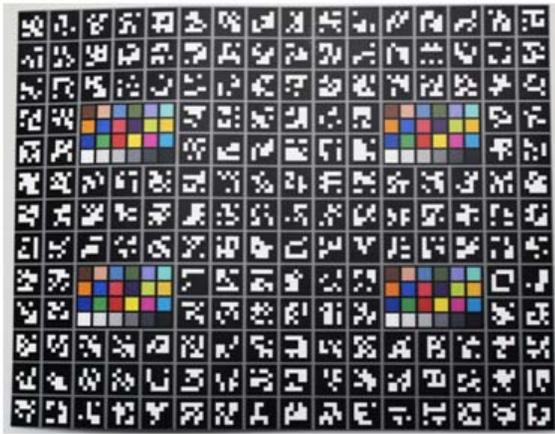


Figura 2. Cartão usado para correções geométricas e de cor.

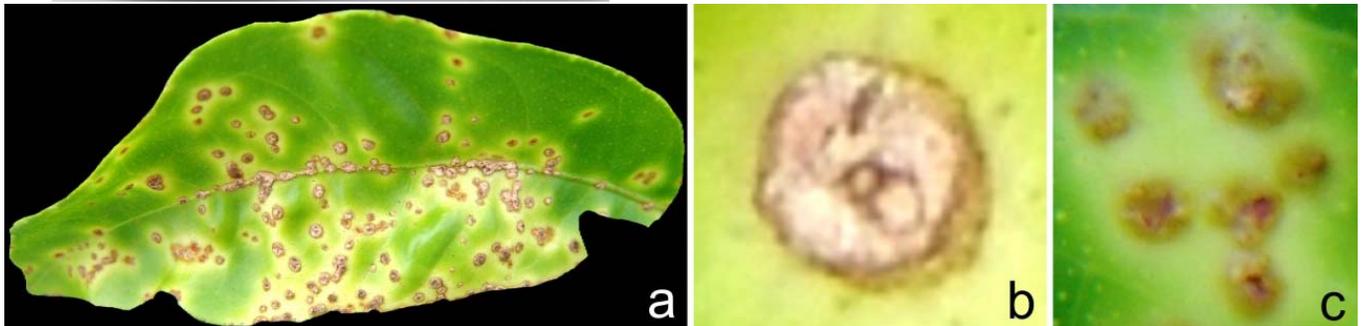


Figura 3. Exemplo de sintomas pequenos espalhados (a), lesões isoladas (b) e aglomerados de lesões (c).



Figura 4. Exemplo de sintomas grandes espalhados (a), lesões espúrias removidas (b) e lesões espúrias mantidas (c).



Figura 5. Exemplo de sintomas isolados (a), região do sintoma delimitada (b) e lesões com duas regiões visíveis (c).

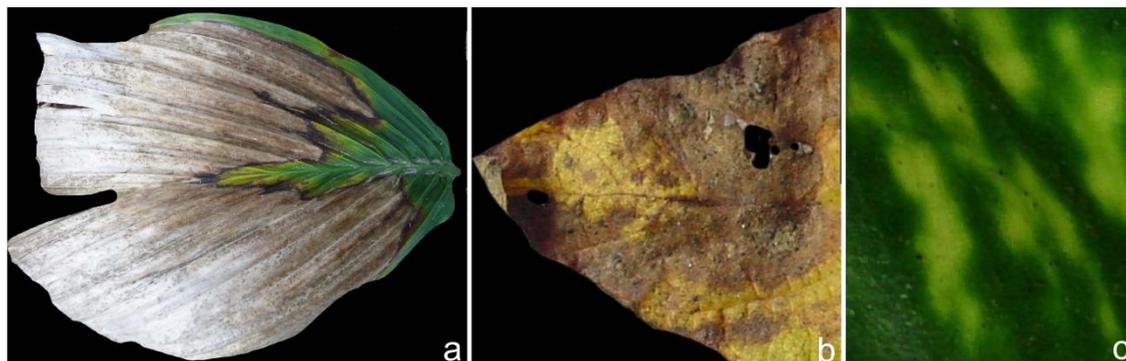


Figura 6. Exemplo de sintoma generalizado (a), lesões delimitadas pela homogeneidade (b) e (c).

Imagens contendo mais que uma doença não foram subdivididas porque a interação entre diferentes doenças pode produzir sintomas inesperados que são difíceis de rotular. Assim, mais estudos são necessários para determinar a melhor maneira de dividir essas imagens. Folhas saudáveis também não foram subdivididas, uma vez que estas não possuem lesões ou manchas.

Em alguns casos, a mesma região pode aparecer em diferentes imagens com diferentes escalas. Isto é particularmente comum no caso de sintomas generalizados, em cujos casos folhas inteiras, além de partes dessas mesmas folhas, são parte da base XDB (Fig. 8).

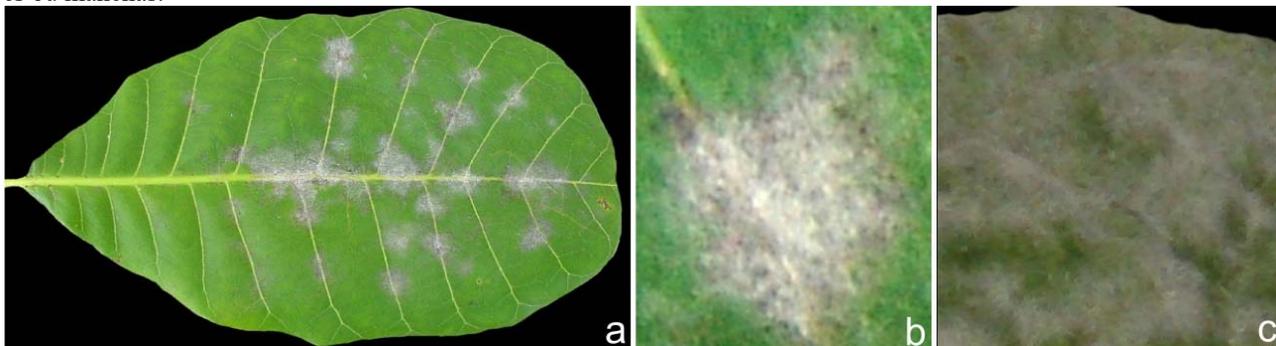


Figura 7. Exemplo de sintomas pulverulentos (a), manchas isoladas (b) e manchas generalizadas (c).

O principal objetivo de subdividir as imagens originais da base PDDB, além de aumentar a base de imagens, foi fragmentar a informação em partes mais bem definidas, permitindo aos métodos de classificação extrair informação relevante de maneira mais eficiente. Porém, a quantidade de informação presente em cada uma dessas imagens está ligada a alguns fatores:

- Número de pixels: normalmente, quanto maior o número de pixels usado para descrever um sintoma, mais detalhes podem ser observados e, conseqüentemente, mais informação pode ser extraída. Isto é particularmente importante no caso de lesões pequenas. Se a resolução da imagem não é suficientemente alta, ou a imagem foi capturada a uma distância muito grande, as lesões podem não ser representadas adequadamente e nenhum esquema de classificação funcionará corretamente (Fig. 9). Apesar de não ser comum, esta situação foi observada algumas vezes na base PDDB, em cujos casos nenhuma subdivisão foi possível. À medida que sensores evoluem, esta situação tende a se tornar menos comum.

- Foco e qualidade óptica: boa resolução espacial é uma condição necessária, mas não suficiente, para garantir que informação suficiente pode ser extraída da imagem. Atributos sutis só serão visíveis e detectáveis pelo computador se a imagem é suficientemente nítida. Este fator depende tanto de um foco apropriado quanto da qualidade óptica das lentes. Como a maioria dos dispositivos para captura de imagens tem qualidade óptica apenas razoável, a perda de nitidez deve ser

compensada por capturas mais próximas e resoluções mais elevadas. A Fig. 10 mostra um exemplo de informação perdida devido à perda de nitidez. Como uma nitidez perfeita é inviável, todas as imagens na base XDB têm algum grau de suavização, de modo que é importante ter em mente que imagens mais nítidas carregarão mais informação, sendo potencialmente mais úteis para caracterizar uma dada doença.

- Reflexões especulares: quando a superfície reflete uma grande parte da luz de volta ao sensor, o brilho excessivo resultante reduz a gama de valores dos pixels naquela região, diminuindo o contraste e, conseqüentemente, causando perda de informação (Fig. 11). Cerca de 15% das imagens na base PDDB têm regiões com reflexões especulares moderadas a severas (reflectância pelo menos 25% maior que regiões livres do efeito), de modo que as lesões ali localizadas carregarão menos informação que aquelas situadas em outras partes da imagem.

XDB tem algumas vantagens além de sua dimensão. Regiões menores tendem a ser mais homogêneas, de modo que fatores tais como variações de sintomas, múltiplas desordens simultâneas e diferenças de iluminação, as quais causam problemas de classificação [13], têm um efeito menos deletério.

A rotulagem manual das imagens, sendo uma tarefa subjetiva, está sujeita a erros. Quando somente algumas poucas imagens estão disponíveis para cada classe, rótulos errados podem ter um grande impacto no processo de treinamento dos algoritmos. Como XDB oferece uma grande quantidade de imagens para a maioria das doenças, o impacto

de rótulos errados será menor.

Testes preliminares usando redes neurais combinadas com *Deep Learning* forneceram algumas indicações das vantagens da base XDB. Em primeiro lugar, foi determinado que apenas doenças contendo mais de 10 representantes poderiam ser consideradas, o que imediatamente reduziu o número de doenças que poderiam ser consideradas para a PDDB. Adicionalmente, considerando apenas as doenças válidas para ambas as bases de imagens, os índices de acerto usando XDB foram entre 7% (soja) e 32% (cajuzeiro) superiores àqueles obtidos usando PDDB. É importante destacar, contudo, que estes são resultados preliminares, uma vez que o algoritmo utilizado ainda deverá ser aperfeiçoado.

#### IV. CONCLUSÕES

Os esforços para construir uma base de imagens fitopatológicas abrangente foram condensados na base de imagens PDDB, a qual contém 2326 imagens de 171 doenças

e outras desordens afetando 21 espécies de plantas. A fim de aumentar a aplicabilidade do material, a maioria das imagens em PDDB foram subdivididas em sub-imagens menores e mais homogêneas, gerando a base de imagens XDB, a qual inclui 46,513 imagens. Ambas as bases estão disponíveis no endereço <https://www.digipathos-rep.cnptia.embrapa.br/>. Seu conteúdo, o qual é acessível após registro, pode ser usado livremente no contexto acadêmico, desde que sem objetivos comerciais. Espera-se com isso contribuir para o progresso da pesquisa em detecção e identificação automática de doenças em plantas. Espera-se também que as bases continuem a crescer continuamente, tornando-se assim ainda mais representativas dos problemas sanitários que acometem produtores no Brasil e no mundo.

#### AGRADECIMENTOS

Este trabalho foi financiado pela Fapesp (proc. 2013/06884-8) e Embrapa (SEG 03.13.00.062.00.00).



Figura 8. Exemplo de uma mesma região presente em imagens separadas; tanto a folha inteira (esquerda) quanto a região dentro do quadro vermelho (direita) são parte de XDB.

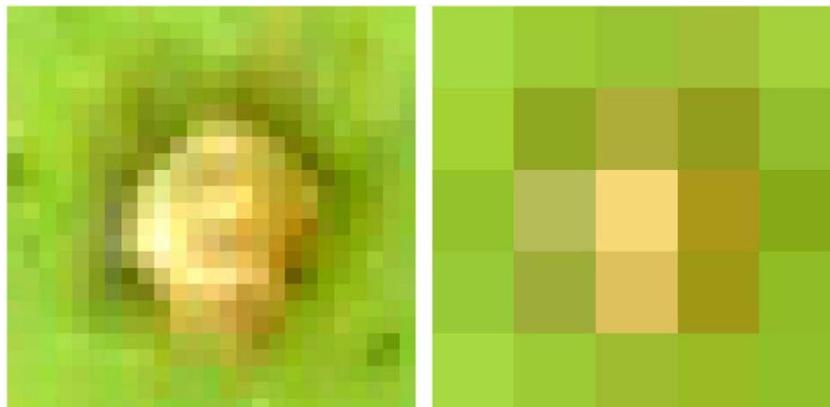


Figura 9. Efeito da resolução da imagem; ambas as imagens mostram a mesma lesão, com a imagem à esquerda tendo 25 vezes mais pixels que a imagem à direita.



Figura 10. Efeito da nitidez; ambas as imagens mostram a mesma lesão, com a imagem à direita tendo sido artificialmente suavizada para similar perda de qualidade.



Figura 11. A imagem à direita está afetada por reflexão especular severa, contendo assim menos informação que a imagem à esquerda.

## REFERÊNCIAS

- [1] R. Prasad, K. R. Ranjan, A. K. Sinha, "AMRAPALIKA: An expert system for the diagnosis of pests, diseases, and disorders in Indian mango," *Knowledge-Based Systems*, vol. 19, pp. 9-21, 2006.
- [2] G. Mansingh, H. Reichgelt, K. M. O. Bryson, K.M.O., "CPEST: An expert system for the management of pests and diseases in the Jamaican coffee industry," *Expert Systems with Applications*, vol. 32, pp. 184-192, 2007.
- [3] R. Oberti, M. Marchi, P. Tirelli, A. Calcante, M. Iriti, A. N. Borghese, "Automatic detection of powdery mildew on grapevine leaves by image analysis: optimal view angle range to increase the sensitivity," *Computers and Electronics in Agriculture*, vol. 104, pp. 1-8, 2014.
- [4] G. Polder, G. W. A. M. van der Heijden, J. van Doorn, T. A. H. M. C. Baltissen, "Automatic detection of tulip breaking virus (TBV) in tulip fields using machine vision", *Biosystems Engineering*, vol. 117, pp. 35-42, 2014.
- [5] A. Pourreza, W. S. Lee, R. Ehsani, J. K. Schueller, E. Raveh, "An optimum method for real-time in-field detection of Huanglongbing disease using a vision sensor", *Computers and Electronics in Agriculture*, vol. 110, pp. 221-232, 2015.
- [6] A. Pourreza, W. S. Lee, E. Etxeberria, A. Banerjee, "An evaluation of a vision-based sensor performance in Huanglongbing disease identification," *Biosystems Engineering*, vol. 130, pp. 13-22, 2015.
- [7] J. Zhang, L. Yuan, R. Pu, R. W. Loraamm, G. Yang, J. Wang, "Comparison between wavelet spectral features and conventional spectral features in detecting yellow rust for winter wheat," *Computers and Electronics in Agriculture*, vol. 100, pp. 79-87, 2014.
- [8] R. Zhou, S. Kaneko, F. Tanaka, M. Kayamori, M. Shimizu, "Disease detection of Cercospora Leaf Spot in sugar beet by robust template matching," *Computers and Electronics in Agriculture*, vol. 108, pp. 58-70, 2014.
- [9] S. Phadikar, J. Sil, A. K. Das, "Rice diseases classification using feature selection and rule generation techniques," *Computers and Electronics in Agriculture*, vol. 90, pp. 76-85, 2013.
- [10] R. Pydipati, T. F. Burks, W. S. Lee, "Identification of citrus disease using color texture features and discriminant analysis," *Computers and Electronics in Agriculture*, vol. 52, pp. 49-59, 2006.
- [11] P. Sanyal, S. C. Patel, "Pattern recognition method to detect two diseases in rice plants," *Imaging Science Journal*, vol. 56, pp. 319-325, 2008.
- [12] J. G. A. Barbedo, L. V. Koenigkan, T. T. Santos, "Identifying multiple plant diseases using digital image processing," *Biosystems Engineering*, vol. 147, pp. 104-116, 2016.
- [13] J. G. A. Barbedo, "A review on the main challenges in automatic plant disease identification based on visible range images," *Biosystems Engineering*, vol. 144, pp. 52-60, 2016.
- [14] J. G. A. Barbedo, "An Automatic Method to Detect and Measure Leaf Disease Symptoms Using Digital Image Processing," *Plant Disease*, vol. 98, pp. 1709-1716, 2014.
- [15] J. G. A. Barbedo, "Identifying Central Features of Cotton Leaves in Digital Images with Difficult Backgrounds," *IEEE Latin America Transactions*, vol. 13, pp. 3072-3079, 2015.
- [16] J. G. A. Barbedo, "A new automatic method for disease symptom segmentation in digital photographs of plant leaves," *European Journal of Plant Pathology*, vol. 147, pp. 1-12, 2016.
- [17] J. G. A. Barbedo, "A novel algorithm for semi-automatic segmentation of plant leaf disease symptoms using digital image processing," *Tropical Plant Pathology*, vol. 41, pp. 210-224, 2016.
- [18] I. Goodfellow, Y. Bengio, A. Courville, *Deep Learning*, Boston, MA. MIT Press, 2016.
- [19] A. Krizhevsky, I. Sutskever, G. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," *Proc. Neural Information and Processing Systems*, 2012.

**Jayme Garcia Arnal Barbedo** atua como pesquisador na Embrapa Informática Agropecuária, em Campinas, São Paulo. Suas áreas de interesse incluem processamento digital de imagens, aprendizado de máquina e aplicação de drones a problemas agropecuários.

**Luciano Vieira Koenigkan** é analista da Empresa Brasileira de Pesquisa Agropecuária. Tem experiência na área de Ciência da Computação, com ênfase em Processamento Digital de Imagens e Sinais, atuando principalmente nos seguintes temas: agricultura de precisão e computação aplicada à agropecuária.

**Bernardo de Almeida Halfeld Vieira** é Pesquisador A da Embrapa Meio Ambiente. Tem experiência na área de Agronomia, com ênfase em Fitopatologia, atuando principalmente nos seguintes temas: fitopatologia, controle biológico, bacteriologia, micologia, *Solanum lycopersicum*, *Citrullus lanatus*, *Acacia mangium*, *Vigna unguiculata*, *Passiflora edulis* e etiologia.

**Rodrigo Veras da Costa** é pesquisador da Embrapa Milho e Sorgo, atuando na área manejo de doenças nas culturas do milho e sorgo.

**Kátia de Lima Nechet** é pesquisadora da Embrapa Meio Ambiente. Atualmente, tem como principais linhas de pesquisa o controle biológico de plantas daninhas com fungos fitopatogênicos; enzimas relacionadas a defesa de frutas em controle alternativo em pós-colheita e etiologia e monitoramento de fungos fitopatogênicos em sistemas de produção.

**Cláudia Vieira Godoy** é pesquisadora da Embrapa Soja, onde atua na área de Fitopatologia com ênfase em epidemiologia e controle de doenças de soja.

**Murillo Lobo Junior** é pesquisador da Embrapa Arroz e Feijão. Tem experiência em manejo integrado de doenças, epidemiologia, ecologia de microrganismos do solo e controle biológico, e na coordenação de projetos financiados pela Embrapa e CNPq.

**Flávia Rodrigues Alves Patrício** é pesquisadora do Instituto Biológico de São Paulo. Tem experiência em doenças do cafeeiro, controle biológico de doenças de plantas, solarização e patologia de sementes.

**Viviane Talamini** é pesquisadora na Embrapa Tabuleiros Costeiros e trabalha com a epidemiologia e o controle de doenças dos citros, da bananeira e do coqueiro.

**Luiz Gonzaga Chitarra** é pesquisador Embrapa Algodão. Tem experiência em Fitopatologia, atuando principalmente nos seguintes temas: *Gossypium hirsutum* L., algodão, manejo de doenças, epidemiologia e fungicidas.

**Saulo Alves Santos de Oliveira** é pesquisador da Embrapa Mandioca e Fruticultura, e desenvolve trabalhos relacionados às principais doenças da

mandioca. Tem experiência na área de Fitopatologia, com ênfase em epidemiologia, biologia de populações de fitopatógenos e resistência de plantas a doenças.

**Alessandra Keiko Nakasone Ishida** é pesquisadora da Embrapa Amazônia Oriental. Tem experiência na área de Agronomia, com ênfase em Fitopatologia, atuando principalmente nos seguintes temas: bacteriologia de plantas, doenças de espécies frutíferas, resistência induzida, controle alternativo e epidemiologia.

**José Maurício Cunha Fernandes** é pesquisador da Embrapa Trigo. Suas pesquisas são focadas em epidemiologia e mais especificamente na modelagem de epidemias de doenças de plantas e em modelos de simulação de culturas.

**Thiago Teixeira Santos** é pesquisador da Embrapa Informática Agropecuária. Tem experiência na área de Ciência da Computação, com ênfase em Processamento Gráfico (Graphics), atuando principalmente nos seguintes temas: fenotipagem baseada em visão computacional, arquitetura de plantas e metrologia baseada em visão computacional.

**Fabio Rossi Cavalcanti** é pesquisador da Embrapa Uva e Vinho. Tem experiência na área de Fisiologia, com ênfase em fisiologia do estresse biótico/abiótico em plantas.

**Daniel Terao** é pesquisador da Embrapa Meio Ambiente, tendo sua pesquisa focada na gestão sustentável de doenças em plantas tropicais pela aplicação de meios de controle físicos, biológicos e alternativos, sem a produção de resíduos tóxicos.

**Francislene Angelotti** é pesquisadora da Embrapa Semiárido. Tem experiência na área de Agronomia, com ênfase em Proteção de Plantas. Atualmente é pesquisadora da Embrapa Semiárido, atuando na área de Mudanças Climáticas Globais, principalmente com monitoramento, avaliação, mitigação e adaptação dos impactos nos sistemas naturais e agrícolas.